UNESP – UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE PRESIDENTE PRUDENTE
FACULDADE DE CIÊNCIAS E TECNOLOGIA
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS CARTOGRÁFICAS

LEANDRO FERNANDES COLADELLO

Integration of heterogeneous data in time series: a study of the evolution of aquatic macrophytes in eutrophic reservoirs based on multispectral images and meteorological data

PRESIDENTE PRUDENTE
2020

LEANDRO FERNANDES COLADELLO

Integration of heterogeneous data in time series: a study of the evolution of aquatic macrophytes in eutrophic reservoirs based on multispectral images and meteorological data

Ph.D. Dissertation thesis presented to Postgraduate Program in Cartographic Sciences of Universidade Estadual Paulista – School of Technology and Sciences – Presidente Prudente campus, as part of requirements to obtain the PhD degree.

Advisor: Prof. Dr. Maria de Lourdes Bueno Trindade Galo

Co-advisors:

Prof. Dr. Ivana Ivánová

Prof. Dr. Milton Hirokazu Shimabukuro

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação - Diretoria Técnica de Biblioteca e Documentação - UNESP, Campus de Presidente Prudente

Coladello, Leandro Fernandes.
C642i Integration of heterogeneous data in time series : a study of the evolution of aquatic macrophytes in eutrophic reservoirs based on multispectral images and meteorological data / Leandro Fernandes Coladello. - 2020
88 f. : il.

Orientadora: Maria de Lourdes Bueno Trindade Galo
Coorientadores: Ivana Ivanová, Milton Hirokazu Shimabukuro
Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente, 2020
Inclui bibliografia

1. Macrófitas aquáticas. 2. Análise de séries temporais. 3. Imagens de sensoriamento remoto. I. Galo, Maria de Lourdes Bueno Trindade. II. Ivanová, Ivana. III. Shimabukuro, Milton Hirozaku. IV. Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente. V. Título.

Alessandra Kuba Oshiro Assunção
CRB-8/9013

unesp **UNIVERSIDADE ESTADUAL PAULISTA**

Câmpus de Presidente Prudente

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Integration of heterogeneous data in time series: a study of the evolution of aquatic macrophytes in eutrophic reservoirs based on multispectral images and meteorological data

AUTOR: LEANDRO FERNANDES COLADELLO
ORIENTADORA: MARIA DE LOURDES BUENO TRINDADE GALO
COORIENTADORA: IVANA IVÁNOVÁ
COORIENTADOR: MILTON HIROKAZU SHIMABUKURO

Aprovado como parte das exigências para obtenção do Título de Doutor em CIÊNCIAS CARTOGRÁFICAS, área: Aquisição, Análise e Representação de Informações Espaciais pela Comissão Examinadora:

Profa. Dra. MARIA DE LOURDES BUENO TRINDADE GALO
Departamento de Cartografia / Faculdade de Ciencias e Tecnologia de Presidente Prudente

Profa. Dra. VILMA MAYUMI TACHIBANA **(participação por videoconferência)**
Departamento de Estatística / Faculdade de Ciências e Tecnologia de Presidente Prudente - UNESP

Profa. Dra. FERNANDA SAYURI YOSHINO WATANABE **(participação por videoconferência)**
Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente

Prof. Dr. JOSEPH AWANGE **(envio de parecer circunstanciado)**
School of Earth and Planetary Sciences / Curtin University

Dr. EDSON EYJI SANO **(participação por videoconferência)**
Centro de Pesquisa Agropecuária dos Cerrados / Empresa Brasileira de Pesquisa Agropecuária

Presidente Prudente, 30 de março de 2020

ACKNOWLEDGMENTS

I'm very grateful to God, my family, my advisors and my friends.

The author also gratefully acknowledges the NASA/USGS for providing Landsat imagery and the NOAA/ESRL for providing climate data, both free of charge. The authors would like to thank the São Paulo State University (Unesp), School of Technology and Sciences, Presidente Prudente, for the infrastructure provided and the R Project for Statistical Computing for making available a free software environment for statistical computing and graphics.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

ABSTRACT

River damming for electric power production usually triggers anthropic activities that strongly impact on aquatic ecosystem. One of the consequences of installing small reservoirs in regions subject to an intense process of urbanization and industrialization is the over-abundance of macrophytes, resulting from the input of nutrients in high concentration into the aquatic ecosystem. Currently, the large volume of multitemporal remote sensing images available in open data sources, as well as the high computational performance that allow the mining of large volumes of data has made the monitoring of environmental phenomena a recurrent object of analysis. The aim of this study is to develop a methodology based on the integration of heterogeneous data, provided by time series of multispectral and multitemporal Landsat images and collections of historical climatic data, to investigate the evolution and spatial behavior of aquatic macrophytes in lakes and eutrophic reservoirs. So, the extensive temporal collection of the Landsat surface reflectance images made available as well as environmental variables data permitted the construction and analysis of time series to investigate the recurrent over-abundance of macrophytes in Salto Grande reservoir, located in the metropolitan region of Campinas, São Paulo, Brazil. Initially, it was found that the the Landsat images have the necessary radiometric quality to perform the time series analyses, through an assessment based on information about the Pixel Quality Assessment Band made available in the metadata of the scenes. Applying Principal Component Analysis to the NDVI time series and using K-Means to cluster the scores of the first component, the reservoir sectors were segmented based on the NDVI temporal variability. Using the BFAST algorithm, it was possible to analyze the persistence of macrophytes in the reservoir. The most abrupt changes, found by the breakpoints, highlighted the importance of monitoring the regions of the dam and near the river. This approach made it possible to map the spatial distribution of macrophytes, to identify periods when abrupt disruptions occurred in the various sectors and to verify the increase in the occurrence of macrophytes in almost the entire reservoir. The influence of exogenous variables on macrophyte infestation of the reservoir was investigated considering two scenarios. Firstly, the effect of the anthropogenic occupation around the reservoir was evaluated comparing NDVI and the NDBI time series. The influence of weather conditions was investigated through joint analysis of NDVI time series and climatic oscillation time profiles, constructed from the NOAA historical data of the air temperature, meridional and zonal winds and precipitable water. With a higher variability, NDVI showed similar seasonal patterns with NDBI while maintaining an increase in trend for all studied period, confirming that the macrophytes in the reservoir still are in greening process. Moreover, this greening process is also highlighted when not only trend and seasonality are correlated in specific locations of the area with the climatic variables, but also when incorporating these three variables to the NDVI series through a Granger-causality model. When combined to lagged values, the climatic variables were found to Granger-causality NDVI, thus, important to describe and predict the macrophytes' persistence.

Keywords: time series analysis, aquatic macrophytes, remote sensing images, spatiotemporal dynamics, NDVI, causality.

RESUMO

O represamento de rios para a produção de energia elétrica usualmente provoca atividades antrópicas que impactam um ecossistema aquático fortemente. Uma das consequências de se instalar pequenos reservatórios em regiões sujeitas à intensos processos de urbanização e industrialização é a abundância de macrófitas, resultante do despejo de nutrientes em grandes concentrações no ecossistema aquático. Recentemente, o grande volume de images multitemporais de sensoriamento remoto disponíveis em bancos de dados gratuitos, bem como a alta performance computacional que permite a mineração de grandes volumes de dados, fazem com que o monitoramento de fenômenos ambientais seja um objeto de estudo recorrente. O propósito desse estudo é desenvolver uma metodologia baseada na integração de dados heterogêneos, fornecidos por séries temporais de coleções de imagens multiespectrais e multitemporais Landsat e coleções de dados climáticos históricos, para investigar a evolução e comportamento espacial de macrófitas aquáticas em lagos e reservatórios eutrofizados. A extensa coleção temporal de imagens de superfície de reflectância Landsat disponível e também dados de variáveis ambientais permitiram a construção e análise de séries temporais para investigar a recorrente abundância de macrófitas no reservatório de Salto Grande, localizado na região metropolitana de Campinas, São Paulo, Brasil. Inicialmente, foi encontrado que as imagens Landsat possuem a qualidade radiométrica necessária para se realizar as análises de séries temporais, através da avaliação baseada na informação contida na banda "Pixel Quality Assessment", disponível como metadado das cenas estudadas. Aplicando-se Análise de Componentes Principais às séries temporais de NDVI e usando K-médias para agrupar os escores do primeiro componente, setores do reservatório foram segmentados baseados na variabilidade temporal do NDVI Usando o algoritmo BFAST, foi possível analisar a persistência das macrófitas no reservatório. As mudanças mais abruptas, encontradas por pontos de quebra, destacaram a importância do monitoramento da região da usina e próxima ao rio. Esse método possibilitou o mapeamento da distribuição espacial das macrófitas, a identificação de períodos de interrupções abruptas ocorridas em vários setores, e a verificação do aumento da ocorrência de macrófitas em praticamente todo o reservatório. A influência de variáveis exógenas na infestação de macrófitas no reservatório foi investigada considerando dois cenários. Primeiramente, o efeito da ocupação antrópica em torno do reservatório foi avaliado comparando as séries temporais de NDVI e NDBI. A influência de condições climáticas foi investigada através da análise conjunta das séries temporais de NDVI e perfis de oscilações climáticas construídos pelos dados históricos NOAA de temperatura do ar, ventos meridionais e zonais, e água precipitável. Com maior variabilidade, NDVI mostrou padrões sazonais similares com o NDBI enquanto mantendo um aumento de tendência por todo o período estudado, confirmando que as macrófitas no reservatório ainda estão em processo de floração. Ainda, esse processo de floração é destacado não apenas através da tendência e sazonalidade correlacionada em localidades específicas com variáveis climáticas, mas também ao se incorporar essas três variáveis climáticas à série de NDVI através de modelos de causalidade de Granger. Quando combinado com valores defasados, as variáveis climáticas foram responsáveis em Granger-causar o NDVI, e, portanto, importante em destacar e predizer a persistência de macrófitas.

Palavras-chave: análise de series temporais, macrófitas aquáticas, imagens de sensoriamento remoto, dinâmica espaço-temporal, NDVI, causalidade.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

6S – Second Simulation of the Sattelite Signal in Solar Spectrum

BIC – Bayesian Information Criterion

BFAST – Breaks for Additive Seasonal and Trend

CPFL – Companhia Paulista de Força e Luz

DN – Digital Number

EMR – Electromagnetic radiation

ERTS – Earth Resource Technology Satellite

ESRL – Earth System Research Laboratory

ETM+ – Enhanced Thematic Mapper Plus

EVI – Enhanced Vegetation Index

GSFLC – Goddard Space Flight Center

GDAL – Geospatial Data Abstraction Library

L1TP – Level 1 Precision and Terrain

LEDAPS – Landsat Ecosystem Disturbance Adaptive Processing System

MEaSUREs – Making Earth System Data Records for Use in Research Environments

NASA – National Aeronautics and Space Administration

NCEP – National Centre for Environmental Prediction

NDVI – Normalized Difference Vegetation Index

NIR – Near Infrared

NDBI – Normalized Difference Built-up Index

NOAA – National Oceanic and Atmospheric Administration

OLI – Operational Land Imager

PCA – Principal Component Analysis

SLC-Off – Scan Line Corrector Off

SWIR – Short wave infrared

TM – Thematic Mapper

QA – Quality Assessment

USGS – United States Geological Survey

TABLE OF CONTENTS

1 INTRODUCTION

Water resources degradation occurs due to natural and cultural factors, mainly related with the accelerated and sometimes predatory occupation of river's wellsprings and margins, lakes and reservoirs, extending beyond their immediate neighborhood. The future use of these resources requires an integrated management, subsidized by the monitoring of anthropic occupation of their surroudings, with the purpose of identifying and integrating the diversity of components which contributes for the degradation of the aquatic environment.

Several researches (Gómez et al., 2016; Mahmoodzadeh, 2007; Jakubauskas et al., 2012) support the premise that remote sensing is an effective alternative for the inland waters system monitoring. The possibility of acquisition of orbital, sinoptic remote sensing in different dates of the year gives information that allows the inference about the status of the aquatic environment, based on optical properties of their constituents. Moreover, the availability of historical series systematically taken by sensors aboard orbital platforms makes possible not only the aquatic system monitoring, but also the development of time-series based models in order to analyze, map and comprehend the degradation processes.

Monitoring natural resources and identifying potential unusual phenomena in a region are possible both in terrestrial and aquatic environments by inspecting spatial, temporal and thematic properties of an image or a set of them (Gómez et al., 2016). Mahmoodzadeh (2007) highlighted that remote sensing sensors are able to scan a region repeatedly, which is necessary to detect changes in regional and global scales. However, the increased availability of computational tools to extract information from multispectral images, such as arithmetic operations among bands to enhance some spectral characteristics of the target or machine learning algorithms, has made the use of remote sensing in mapping and monitoring natural resources recurrent.

The great volume of multispectral images (in terms of number of images and number of bands per image) taken by the United States (U.S.) satellites' sensors of Landsat program, available since 1972, Multispectral Scanner (MSS – Landsat 1), Landsat 5 (Thematic Mapper – TM), Landsat 7 (Enhanced Thematic Mapper Plus – ETM+), and the more recent Landsat 8 (Operational Land Imager – OLI) in 2013, allow significant amount of time series generation based on information provided by these images. As reported by França et al. (2007), availability of these images permits the access of information from extensive areas in regular time periods. Even with this use of remotely sensed products, in contrast, small water bodies like artificial reservoirs still receives little attention from the research community, while large ones are regularly monitored (e.g., Zhang et al., 2016). According to Jones et al. (2017), this may happen because information from small areas like reservoirs is challenging to compile and to update at regional or global levels, unless spatial and temporal resolutions of the satellite images are sufficiently high.

It is possible to integrate data from other sources to the spectral bands of original multispectral sensors or those resulting from some processing that enhances some important characteristics, which together can contribute to the analysis of the relation of the physical environment and the human occupation, for example. Meteorological variables as temperature, precipitation, direction and wind speed, relative air humidity and others are also available in temporal collections and can be considered when using appropriate statistical tools as a way to integrate this ancillary information in the characterization of temporal changes (Lu et al., 2004).

The substantial volume of data available in multitemporal imageries acquired from TM, ETM+ and OLI multispectral sensors as well as historical series of meteorological variables create the possibility of the joint use of heterogeneous data in water resources monitoring like small reservoirs. At the same time, the growing capacity and models for data processing in different formats and origins, is in synchrony with the development of new software tools, enabling the construction of several scenarios to effectively explore the evolution of phenomena that occur in aquatic and terrestrial environments.

The hypothesis of this research is that the combination of different scenarios, built by integrating remote sensing multitemporal data and environmental variables provided by other sources in time series configuration, permit to monitor the evolution of specific environmental phenomena and to identify natural or culturally significant events that provoked ruptures in eutrophicated water bodies.

To test this hypothesis, it was identified that high aquatic macrophytes infestations is recurrent in the Salto Grande reservoir, formed by the impoundment of Atibaia river during the installation of Americana’s hydroelectric plant in 1949. Inserted in the metropolitan region of Campinas, the reservoir’s surroundings suffered an intense anthropic occupation process, starting with townhouse and leisures area constructions and spreading progressively with the urbanization and industrialization of the region until the massive development of sugarcane plantations. The landscape’s surrounding degradation was accompanied by degradation in the water quality of the reservoir, which accelerated the development of extensive banks of aquatic emerged macrophytes accumulated on the margins and in the dam’s vicinity.

1.1 Objectives

The main aim of this research is to develop a methodology based on hetereogeneous data integration – provided by the time series of multispectral and multitemporal Landsat images and collections of historical climatic data – to investigate the evolution and spatial behavior of aquatic macrophytes in eutrophic lakes and reservoirs. The specific objectives are:

- assess the radiometric quality of the fragments of the Landsat images to investigate, over time, the occurrence of aquatic macrophytes in a water body and the influence of its surroundings, based on information about the Pixel Quality Assessment Band made available in the metadata of the scenes;
- explore the potential of Landsat image-attributes' time series to monitor the spatiotemporal behavior of aquatic macrophytes, enabling the analysis of trends and changes in the macrophytes occurrence in reservoirs;
- investigate whether the dispersion of macrophytes banks is influenced by climatic conditions – exogenous factors – expressed by climatic variables such as temperature, precipitation, wind direction, beyond the land use (e. g., anthropic actions) on its surroundings.

1.2 Motivation

The extensive historical imagery collection made available by sensors aboard orbital platforms has been object of study in remote sensing due to the recent computational hardware and software improvements, enabling not only the development of robust methodologies but also the operational big data analysis. Using multisource data, naturally the question of data integration emerges and, according to Tolpekin and Stein (2013), it is useful in monitoring processes and changes, mainly when time series are used. While it is possible to integrate the sensor information with other data sources, such as elevation models, land use and land cover, field measurements, forecasting models, non-spatial data, etc., it is essential to seek temporal information to adhere this existing information. Jakubauskas et al. (2002) showed that the time domain added to multispectral data often provides more information about the condition of vegetation than spatial information, for example.

The unsustainable blooming of aquatic vegetation changes the relationship between natural vegetation and climatic conditions producing significative changes in the water and ground condition (Jakubauskas et al., 2002; Londe, 2016). Investigation and identification of general and seasonal patterns in the aquatic macrophytes behavior, according to its spreading, dispersion and movement through the reservoir, as well as its relation with the anthropic occupation in the reservoir surroundings, need more than a static representation of extracted information in multispectral images. Thus, it is necessary to integrate datasets from other sources and highlight their temporal perspective. This condition is common in many reservoirs located in areas with a high degree of human occupation, thus, the methodology proposed in this research can constitute a reference to define strategies of management and control of infestation by aquatic macrophytes in reservoirs. Moreover, according to Eckert et al. (2015**)** and Liu et al. (2018)**,** there is a lack of studies that combine climatic variables and the Normalized Difference Vegetation Index (NDVI) in long-term time series analysis in order to assess the stability of temporal changes, and, there is a lack of research about small reservoir areas.

To achieve the objectives proposed here, not only NDVI is studied together with climatic variables, as perturbations from them are observed in this index, but also the correlation among these variables is addressed to analyse the response of aquatic vegetation (Jakubauskas et al., 2012; Luo et al., 2016). Another metrics, the Normalized Difference Built-Up Index (NDBI) is used to assess the imperviousness of the surrounding areas (Zha, Gao & Ni, 2003). These time series indices are decomposed by using the Breaks for Additive Season and Trend (BFAST) algorithm (Verbesselt et al., 2010) to inspect their temporal components. Villa et al. (2018) pointed out that consistent satellite data can be used to monitor and quantify intra and interannual trends in vegetation cycles. The BFAST algorithm is also suitable to be applied in these time series indices as standard time series decomposition methods that are not directly applicable to identify changes (Verbesselt et al., 2010). Finally, causality models using the exogenous variables are developed to investigate if their values have some potential causal association with the NDVI. Granger causality test, for example, has been used in studies like climatology and atmosphere conditions (Lee et al., 2015).

The question of quality of data is also important: the data obtained by the government, such as meteorological data can be inspected through the metadata available in the site directly, but a small portion of multispectral images may not reflect the metadata available for an entire image. Landsat 8 images, for example, have a pixel resolution of 30m for visible bands and have a scene size of 170 km x 185 km (USGS, 2019). From this perspective, based on obstructions present in an image, an overall indicator of quality for the entire image may not reflect the quality of a particular small area. For example, cluster of obstructions still yields good parts of an image, while a scattering of obstructions might make the image worthless (USGS, 2018b).

Finally, the methodology proposed here for the monitoring of eutrophic reservoirs makes available not only a concrete process for building time series for these water bodies analysis, but also an indicator of the quality of the images and a metric can be incorporated in time series models to guide decision makers about the feasibility of using the extracted information. Moreover, the scenario explored in the case study is expected to be reproduceable in other small reservoirs by using the statistical tools presented here, with the addition of variables like climatic ones.

1.3 Thesis structure

This thesis is structured in five sections that describe the research developed. Section 1 introduced the problem and presented the objectives and motivation Section 2 presented the general theoretical background that embodied the research. In Section 3, the methodologies for experiments 1, 2 and 3 are described. Section 4 shows all results and discussion achieved by each one of the experiments. In Section 5, the concluding remarks and next steps are addressed.

The above structure was adopted to suit to the three main papers being written based on the methodology developed and experiments carried out. One entitled “Macrophyte’s abundance changes in eutrophicated tropical reservoirs exemplified by Salto Grande (Brazil): Trends and temporal analysis exploiting Landsat remotely sensed data” was already submitted to the Journal of Applied Geography. The others are currently being concluded to be sent to specific journals and are entitled “Analyzing the influence of exogenous factors on the spatiotemporal dynamics of aquatic macrophytes by time series from spectral indices and climate data” and “Assessing the quality of images in time series to analyze the spatiotemporal dynamics of aquatic macrophytes in small eutrophic reservoirs”.

2 THEORETICAL BACKGROUND

This section is designed to explain the general theory used in majority of this study and supports the proposed methodology. Specific topics which appear in the experiments realized are treated in their corresponding chapters.

2.1 Multispectral images and spectral indices

Multispectral images of terrestrial surface are obtained by sensors on board the satellites or aircrafts by the detection and measurement of interaction of electromagnetic radiation (EMR) with terrestrial materials, notably at wavelengths in which radiation is reflected.

Among the systems currently in operation, there is the U.S. Landsat American series. Started as Earth Resource Technology Satellite (ERTS) in 1972, several more modern satellites were launched with more sophisticated sensors on board. The Thematic Mapper (TM) sensor, on board the Landsat 4 and Landsat 5 satellites, had seven spectral bands, of which six operated in optical spectrum with 30 meters spatial resolution. The last band was a thermal one with 120 meters resolution. The Enhanced Thematic Mapper Plus (ETM+), on board the Landsat 7 and launched in 1999, contains the same spectral bands of the previous satellites, however, a panchromatic band was added with 15 meters spatial resolution. In 2013, on board the Landsat 8, the Operational Land Imager (OLI) sensor also contains similar characteristics of TM sensor, with the addition of a coastal (ultra-blue) band and it is useful to detect cirrus-type clouds.

Considering multispectral images, a pixel is referred as the position of an element of resolution in terrestrial surface containing information that is provided as a digital number (DN). The DN is the quantization of the irradiance integrated in each pixel transformed in electric signal. This number belongs to the $[0,\ 2^Q - 1]$ interval, with Q being a finite number of bits, which can change according to the radiometric resolution of the sensor. The higher values of Q, the closer the quantized value will approach the sensor's generated signals (Schowengerdt, 2006). As an example, sensors installed on board the Landsat 5 and 7 satellites have radiometric resolution – Q – of 8 bits, whereas Landsat 8 sensors have 16 bits resolution, but only 12 bits are used for quantization.

However, the DN value does not have a physical meaning and it does has limitations if the targets from the images are not specific (Chavez Jr., 1988). To detect surface changes, results must be related to physical ground units. In general, methods to convert DN into a physical unit, e.g., reflectance (Milton, 1987, for example) are based on sensor's gains and offsets, solar irradiance and zenithal solar angle (Chavez Jr., 1996). The measurement, however, still will contain atmospheric effects as scattering and absorption, and, to remove (or at least minimize) them, some techniques must be applied, as, if not taken in consideration, the reflectance obtained from the DN will be the apparent reflectance and not the surface reflectance. Examples of atmospheric effect removing techniques include the dark object subtraction, which uses the DN directly (Chavez Jr., 1988) and radiative transfer models, which are models implemented in many softwares of image processing, e.g., the Moderate Spectral Resolution Atmospheric Transmittance Algorithm MODTRAM and Second Simulation of the Satellite Signal in the Solar Spectrum (6S) (Kotchenova et al., 2006).

After removing the effects of atmosphere and with the purpose of highlighting some specific kind of information acquired from the multispectral images, arithmetic operations are performed using the different spectral bands, which result in multispectral indices. The reflectance measured is converted into multispectral indices (Myeni et al, 1995). According to Verstraete & Pinty (1996), an index is considered optimal if it is sensitive to the interested target (e.g., vegetation presence), and insensitive to potential perturbation (e.g., atmospheric effects and soil color changes). Vegetation indices are considered as efficient empirical methods because, firstly, they explore the fact that green vegetation can absorb much of the solar radiation in the visible spectrum (400 nm - 700 nm), particularly in blue wavelengths (430 nm), red (660 nm) for α-chlorophyll, and blue (450 nm), red (650 nm) for β-chlorophyll and scatter them at near infrared (NIR). The use of these indices is explained by the fact that a huge volume of data provided by satellites can be processed with low computational effort, enabling spatial and temporal analysis in large scales (Myeni et al, 1995).

From the surface reflectance obtained from red (ρ_{RED}) and NIR (ρ_{NIR}) bands of multispectral images, a metric which enhances vegetation areas of a scene is known as Normalized Difference Vegetation Index (NDVI), given by:

$$NDVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED})} \quad (1)$$

This index results in values within the interval [-1, 1] and is based on the chlorophyll absorption in the red spectral region and the scattering of EMR by the internal structure of the leaves in the NIR interval. In this way, in the absence of any ancillary information, NDVI interpretation is given by the higher the index value, the higher is the probability that the inspected area is covered by green vegetation (Verstraete and Pinty, 1996).

Differently from vegetation areas, which are more homogeneous, urban areas present different land use and land cover classes, e. g., residential areas with distinct densities and socioeconomic structures, commercial and industrial areas, presence of coverings like roofs, highways, streets and others (Herold et al., 2004). Despite being less sensitive to temporal variations (except in emergencies cases like accidents, fires, etc.), the diversity of targets exhibits certain difficult of being classified (Jensen, 2000).

Just like the vegetation areas, built-up areas, aside their urban complexity (Badlani et al., 2017), can also be identified by specific multispectral indices. Zha et al. (2003) proposed arithmetic manipulations in order to obtain an index for built-up areas. This formulation corresponds to the normalized difference between the reflectances of shortwave infrared (ρ_{SWIR}) and NIR (ρ_{NIR}) bands, and is given by:

$$NDBI = \frac{(\rho_{SWIR} - \rho_{NIR})}{(\rho_{SWIR} + \rho_{NIR})} \tag{2}$$

and it is named as Normalized Difference Built-up Index (NDBI). NDBI also has values within the [-1, 1] interval. Positive values mean that the probability to find a a built-up area is high.

Supposing that we have a collection of multispectral images, each image with n bands, it is possible to track a given process by inspecting the pixel values or index values through that collection, for example, using a time series approach.

2.2 Time series of images

Considering the fact that an image can be seen as a matrix, each pixel contains an information on some attribute at a position (x, y). Extracting this information from a determined pixel for each one of the images, it is generated a time series for that specific pixel (Jönsson, 2004). Figure 1 illustrates this situation.

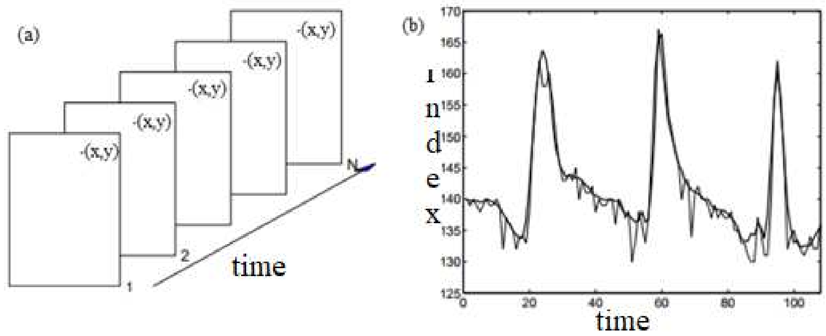


Figure 1. Pixel-wise image time series. (a) Sequence of images in different times with a pixel at position (x,y) highlighted. (b) Time series from the pixel at position (x,y). Source: Adapted from Jönsson (2004).

From the time series obtained for the pixel (x, y), it is possible to study how the evolution of the information presented in the pixel is over time. Let $T = \{t: t_1 < t < t_2\} = \{1, 2, \dots, N\}$ a set of integer numbers, then a time series can be denoted as $\{Y(t): t \in T\}$ and be continuous if its observations are collected continuously over time or be discrete if collected in equispaced time points. The time series $Y(t)$ can be decomposed in three components: trend, seasonality and error, the latter being the part which is not explained neither by trend nor seasonality.

Among the temporal decomposition techniques, there is the Breaks for Additive Seasonal and Trend (BFAST) algorithm, which decomposes a time series into another three series characterized by trend, seasonality and error, through an additive model given by:

$$Y_t = T_t + S_t + e_t, t = 1, \dots, N \qquad (3)$$

in which Y_t is the observation at a time t, T_t is the trend component, S_t is the seasonality component, and e_t is the error, sometimes called also noise of the model.

The BFAST algorithm was proposed by Verbesselt et al. (2010) and is based on the principle that the estimates of changes are a combination of seasonal changes, gradual and abrupt, being the latter caused by events as deforestations, urbanizations, floodings and fires (Verbesselt et al., 2010b). The BFAST algorithm decomposes the time series in the search of gradual and abrupt changes in the seasonal and trend components, being possible of observation by looking for breakpoints. Specifically, a $Y_t = y_t$ observation for some time t is modelled as a linear trend component and a harmonic seasonality:

$$Y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^{k} \gamma_j sin\left(\frac{2\pi jt}{f} + \delta_j\right) + e_t \qquad (4)$$

with unknown parameters being the intercept α_1, trend α_2, amplitudes γ_j, and phases (seasonality) δ_j, $j = 1, \dots, k$. The f term represents the frequency of the time series and e_t represents the error for a time t, with standard deviation σ.

The number and positions of breaks in a time series are unknown and have to be estimated. The parameter h that specifies the minimum number of observations in a segment must be provided so that the maximum number of breaks will be the ratio between this parameter and the number of observations in time (*h/N*) (Watts & Laffan, 2013). The set of breakpoints positions $(b_1, \dots b_m)$ is obtained by a grid search procedure and it is optimum when:

$$\left(\widehat{b_1}, \dots \widehat{b_m}\right) = \arg min_{\left(\widehat{b_1}, \dots \widehat{b_m}\right)} \sum_{i=1}^{N} e_i^2 \qquad (5)$$

i.e., minimizing the sum of squared residuals. The optimal number of breakpoints is given by the minimum Bayesian Information Criterion (BIC) (Gideon, 1978), expressed by:

$$BIC = -2\ln(L) + Wlog(N) \qquad (6)$$

where L is the likelihood for the model and W is the number of parameters and is added as a penalty for the model estimation.

Figure 2 presents a flowchart exemplifying the algorithm procedure, applied in some phenological study.

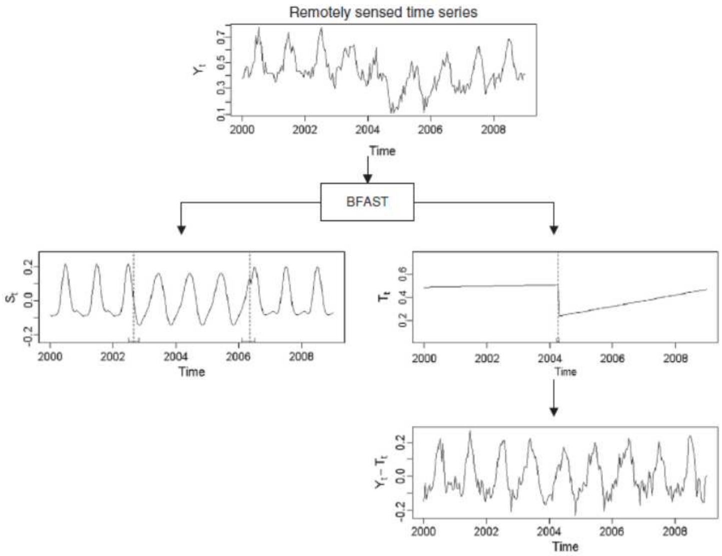


Figure 2. Result obtained from BFAST algorithm (example). Time series of NDVI values is decomposed in seasonality (left) and trend (right). Below the detrended time series is given as an example of the algorithm output. Source: Verbesselt et al., 2010.

In the example of Figure 2, Y_t is the time series representing NDVI values. The BFAST method detects changes in the original time series: two in the seasonal component S_t and one in the trend component T_t. In T_t, it is detected an abrupt change followed by a gradual recovery, as shown by the slope of the line in the right panel. The example also shows the possibility of the time series to be correct by the trend effects, as shown in the last panel by a series $Y_t - T_t$. According to Verbesselt et al. (2010), this procedure can eliminate errors provided by satellite sensors (e. g., calibration problems), which can affect a long-term time series. However, the trend component is important to characterize change phenomena. Although other methods are being used in time series decomposition and analysis, e.g. multivariate linear regresseion analysis (Rieser et al., 2010), BFAST was chosen as it is a method which simultaneously accounts for variation at seasonal scale, while detecting changes in long term trends (gradual interannual and abrubpt intraannual changes) that can be masked by seasonal variability. It also has fast and requires minimum amount of processing time and does not require definition of thresholds (Williams, 2013; Watts and Laffan, 2014). The entire BFAST algorithm is implemented in R language through the package *bfast* (R Development Core Team, 2016).

2.3 Data integration and image quality

Data integration (or data fusion) is the name given to the process of adding different information layers extracted from different sources, spatial or not, in order to synthesize them and to extract new and significant information (Wiemann and Bernard, 2015; Tolpekin and Stein, 2013).

According to Tolpekin and Stein (2013), there are five data categories that can present properties of interest: (1) primary observables from remote sensing, e. g., reflectance, radiance; (2) external variables that influence remote sensing observables, e.g., atmosphere parameters and data acquisition geometry; (3) surface properties that influence remote sensing observables, but they are not subject of interest for a specific study; (4) surface properties, but subject of interest for a specific study; and (5) variables undetectable from remote sensing, but important for the subject.

Wiemann and Bernard (2015) synthesized a plethora of contexts which data integration are used, being based on application field, automation level, operation frequency, data schema, used models, and spatiotemporal orientation. Figure 3 summarizes these contexts in case of spatial data are multispectral images.

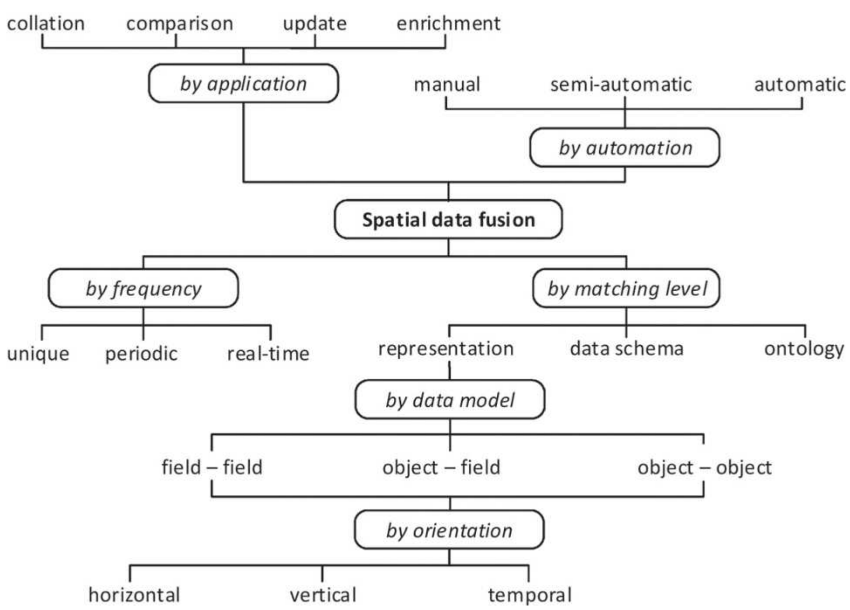


Figure 3. Different contexts of data integration use. Source: Wiemann & Bernard (2015).

From the perspective of this research, spatial data is considered as the information obtained by multispectral images, which was used according to a temporal orientation, in regular periodic times. As non-spatial data, here the environmental variables collected are classified according to Tolpekin and Stein (2013) as information unregistered by the remote sensor, but it is important for the study.

Multispectral Landsat imagery has been providing ways to select a suitable image for scientific analysis. Firstly, USGS provides collection 1 tier 1 Landsat imagery, whose per-pixel quality data and per-pixel cloud masks, radiometric and geometric calibration, and metadata are improved (Dwyer, 2018). Secondly, a general quality score for an entire scene is given in its respective metadata file since 2018, instead of a quality indicator for each one of the available bands of the image (USGS, 2018a). It is important to emphasize that this score refers to quality as the quantity of bad scans or dropped frames encountered in the entire scene and it is given by a integer ranging from 0 (high number of problems) to 9 (no problems) (USGS, 2018b; 2019a, 2019b).

USGS also provides a Quality Assessment (QA) band, which is one possibility to evaluate the suitability of an image for analysis based on pixel information. This band contains for each pixel a decimal value that represents bit-packed combinations of surface, atmosphere and sensor conditions, which can alter the usefulness of a pixel (USGS, 2019a). Here, quality refers to a combination of criteria, and not dropped or bad frames as previously. The QA band is the result of a procedure based on the Fmask algorithm (Zhu and Woodcock, 2014), which was originally designed to provide cloud, cloud shadow and snow occurrence information for Landsat 4 and 7. The algorithm was later improved for Landsat 8 and Sentinel 2, with inclusion of the cirrus band, which provides cirrus cloud occurrences (Zhu et al., 2015).

3 MATERIALS AND METHODS

3.1 Study area and datasets

The Salto Grande reservoir, with a length of 17 km and a perimeter of 64 km, is formed by the damming of the Atibaia River, when the Americana hydroelectric plant, which began its operation in 1949, was installed. Its flooded area varies between 10.55 km² and 13.25 km² (Fonseca, 2014), with a mean depth of 8 m and mean time of retention of 30 days. The Americana hydroelectric plant has a capacity of generating 30 MW, but its operating system has been hampered by the high concentrations of macrophytes, which reached critical values in the last decade. The reservoir is located in three municipalities in the metropolitan region of Campinas (Americana, Paulinia and Nova Odessa), São Paulo, Brazil (Figure 4), in an area characterized by high rates of urbanization and industrialization.

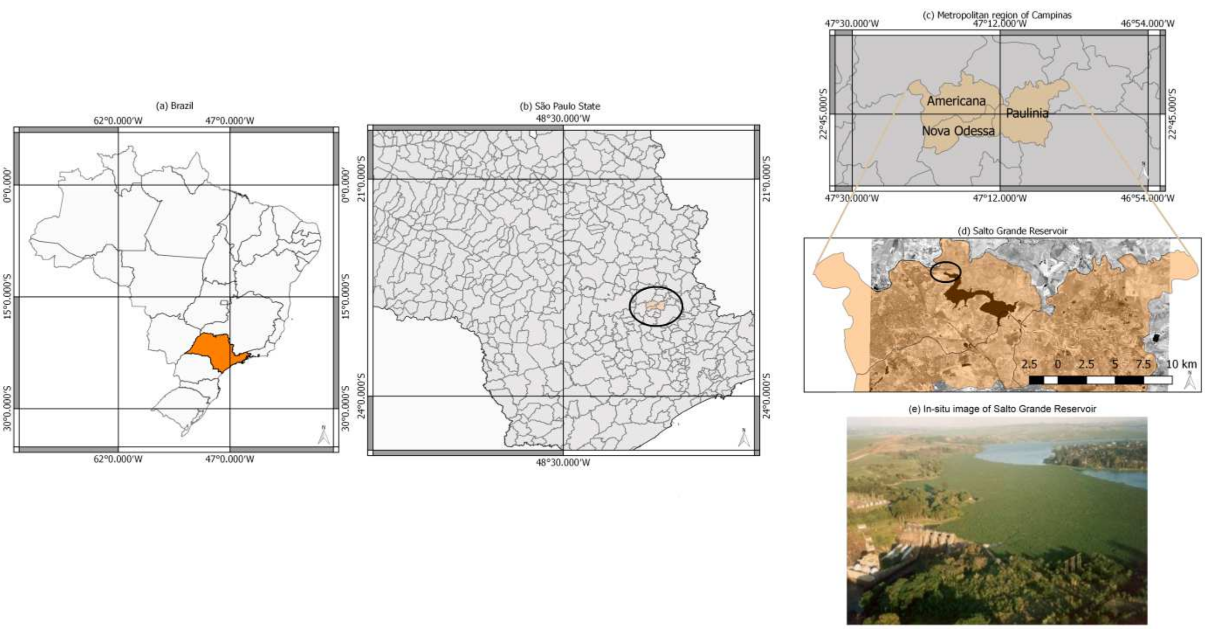


Figure 4. Salto Grande (a) location in Brazil, (b) location in the São Paulo State, (c) location in metropolitan region of Campinas, (d) the reservoir (in black) in a near infrared spectral band of a Landsat image, delineations (in orange) of the municipalities of Americana, Paulinia and Nova Odessa, and (e) aerial image of Salto Grande reservoir showing part of the water body with high infestation by aquatic macrophytes.

Fonseca (2013) reported that the reservoir formation accelerated the accommodation of summer farms and subdivisions of its surroundings, intensified in the 1970 and 1980 decades, with rapid urbanization and population leisure needs. The forward intense anthropic occupation around the reservoir triggered the degradation of the landscape and the decrease of water quality, mainly due to the increase in sugarcane plantations, pasture and domestic pollutant load from nearby cities (CPFL, 2009). A land use and land cover mapping of Salto Grande reservoir was made by Fonseca (2013) and illustrates the current scenario, resulting from this anthropic intervention (Figure 5).

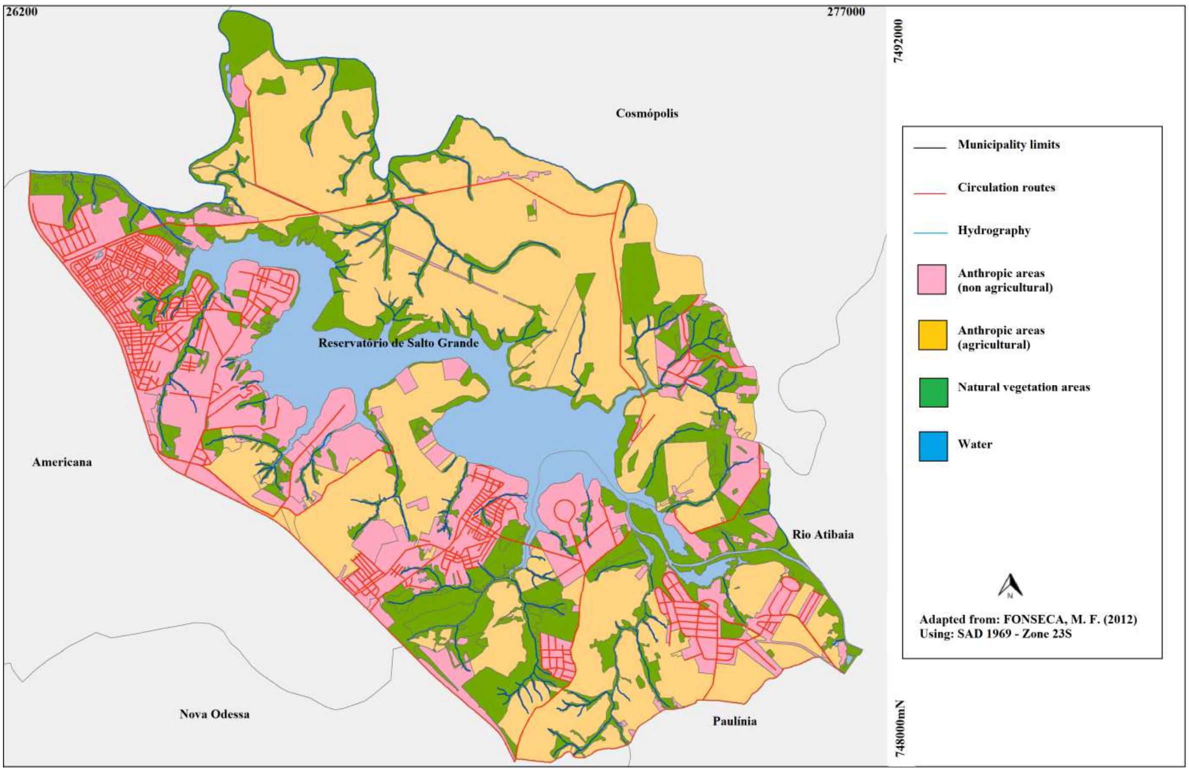


Figure 5. Land use and land cover map from Salto Grande region. Source: Adapted from Fonseca (2013).

The delimitation of the area of influence of the Salto Grande reservoir defined by Fonseca (2013) for the mapping of land use and cover shown in Figure 5 was adopted in this research to create the mask that represents the region surrounding the reservoir.

3.1.1 Landsat multispectral images

Computational routines were developed in Python programming languages (Van Rossum, 1995) and *R* software environment (R Development Team, 2016). For both, existing packages were used for programming, whose routines handled mainly (I) the multispectral images taken by the TM, ETM+ and OLI sensors for the first and third experiments, and (II) climatic data, as well as the multispectral images for the second experiment. The surface reflectance images were downloaded from the *LSDS Science Research and Development* site, *https://espa.cr.usgs.gov/*, of *United States Geology Survey* (USGS) using the reference period of 1984 (earliest available) to 2017.

Landsat imageries made available by USGS, since 2016, are being reorganized into a data collection structure and allocated into tiers 1, 2 and real-time. These classifications were adopted with the intention of organizing the extensive image collection available from sensors on board the Landsat satellites, according to their obtaining method and quality standards, in order to facilitate the correct staking of images and time series analysis.

Tier 1 images have a quality level of geometry and radiometry that contemplate radiometrically calibrated and orthorectified scenes using control points and digital elevation models for topographic variation correction. Images that unfulfill these requirements are classified into Tier 2, which are the ones that did not obtain acceptable cloud cover and/or sufficient control points for their registration, among other reasons. Real-time images are the ones that are available immediately for download using predictive ephemerides. When they are processed with definitive ephemerides, they are classified into Tier 1 or Tier 2.

Surface reflectance images, derived from images of Tier 1, are characterized by USGS as higher quality level products and are available for users after processing using specialized apps, which differentiate themselves by their original satellites. For Landsat 5 and 7, images are processed by the Landsat Ecosystem Disturbance Adaptive Processing System software (LEDAPS), created by the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs), and subsidized by Goddard Space Flight Center (GSFC), and University of Maryland. For Landsat 8, it is used the Landsat 8 Surface Reflectance Code (LaSRC), documented by Vermote et al. (2016). The difference between the two algorithms is reported in the user guide from USGS for the surface reflectance products. The radiative models used by each software are as follows: while LEDAPS uses an adapted version of 6S model, in which the atmosphere is divided in layers and the radiative transfer is numerically calculated for each layer (Kotchenova et al., 2006), LaSRC uses an unique intern algorithm which benefits from the narrow spectral bands of OLI sensor and its aerosol band (Vermote et al., 2016). Because of the quality of Tier 1 images, it was requested for this research, from the study period of 1984 to 2017:

- Multispectral Tier 1 images taken by TM sensor, path/row 219/76, from 29/06/1984 to 28/09/2011, corrected for surface reflectance by USGS. The available images are from bands 1, 2, 3, 4, 5 and 7, with spatial resolution of 30 m, spectral configuration between 450nm and 2350nm, and temporal resolution of 16 days. Finally, the radiometric resolution is 8 bits for the sensor.

- Multispectral Tier 1 images taken by ETM+, path/row 219/76, from 02/08/1999 to 03/08/2017, corrected for surface reflectance by USGS. The available images are from bands 1, 2, 3, 4, 5 and 7, with spatial resolution of 30m, temporal resolution of 16 days and radiometric resolution of 8 bits.

- Multispectral Tier 1 images taken from OLI sensor, also path/row 219/76, from 26/04/2013 to 11/08/2017, corrected for surface reflectance from USGS. The available images correspond to bands 1 to 7, with ground sampling distance of 30m, spectral configuration of 433 to 2290nm, radiometric resolution amplified to 12 bits, thus getting less variation signals registration. Temporal resolution is also 16 days. The characteristics of the bands of each sensor are summarized in Table 1.

Table 1. Spectral configuration of TM, ETM and OLI sensors, on board of Landsat 5, 7 and 8 satellites, respectively.

Bands	Landsat 5 - TM		Landsat 7 – ETM+		Landsat 8 – OLI	
	Spectral Interval (μm)					
1	Blue	0.45 – 0.52	Blue	0.45 – 0.52	Ultra Blue	0.435 – 0.451
2	Green	0.52 – 0.60	Green	0.52 – 0.60	Blue	0.452 – 0.512
3	Red	0.63 – 0.69	Red	0.63 – 0.69	Green	0.533 – 0.590
4	NIR	0.76 – 0.90	NIR	0.77 – 0.90	Red	0.636 – 0.673
5	SWIR1	1.55 – 1.75	SWIR1	1.55 – 1.75	NIR	0.851 – 0.879
6					SWIR1	1.566 – 1.651
7	SWIR2	2.08 – 2.35	SWIR2	2.08 – 2.35	SWIR2	2.107 – 2.294

Source: Adapted from USGS.

The products contain metadata file about the images. They describe the characteristics such as scene identifier, satellite that captured the image, sensor, path/row, acquisition date, etc. Figure 6 shows a piece of a ".xml" archive referring to the band 1 from a TM/Landsat 5, from 28/09/2011.

```
<short_name>LT05SR</short_name>
<long_name>band 1 surface reflectance</long_name>
<file_name>LT05_L1TP_219076_20110928_20161005_01_T1_sr_band1.tif</file_name>
<pixel_size x="30" y="30" units="meters"/>
<resample_method>none</resample_method>
<data_units>reflectance</data_units>
<valid_range min="-2000.000000" max="16000.000000"/>
<app_version>LEDAPS_3.2.1</app_version>
<production_date>2017-10-07T21:10:00Z</production_date>
```


Figure 6. Piece of metadata from a surface reflectance image, band 1, TM/Landsat 5, from 28/09/2011. Information given is the pixel size and the valid values interval.

Download of candidate scenes to compose the time series study was done through the site *https://espa.cr.usgs.gov/*. By using a text archive where the name of each image (for the USGS convention of metadata see Figure 6) is written, it was possible to send a requisition to the site through a Bulk Order. The option enables the processing (e.g., transformation to surface reflectance images) and download of many images at once.

Together with the text file, it is necessary to select the products of interest. For this research we required the surface reflectance images in GeoTIFF format, as well as the metadata for original and transformed images. In a first screening for the Tier 1 selection, only the ones classified by USGS as “level 1 precision and terrain” (L1TP) were selected, as they have the best quality for time series analysis. As mentioned before, they are radiometrically calibrated and orthorectified using control points and digital elevation models for corrections of relief.

Although the image ordering can be done in bulk, to download them, it was necessary to use the Python algorithm named ESPA Bulker Download Master, also created and made available by USGS in *https://github.com/USGS-EROS/espa-bulk-downloader*. It was downloaded 315 TM/Landsat 5 images (29/06/1984 to 28/09/2011), 294 ETM+/Landsat 7 images (02/08/1999 to 03/08/2017) and 88 OLI/Landsat 8 images (26/04/2013 to 15/11/2017), totalizing 697 Tier 1 “L1TP” surface reflectance images. In volume, these images corresponded to 560GB for TM images, 608GB for ETM+ images and 277GB for OLI images.

3.1.2 Climatic variables

According to Lacoul and Freedman (2014), the distribution and abundance of macrophytes are influenced by variations of environmental factors, including flow velocity, irradiance, salinity, ice cover, temperature, nutrients and pollutants. Based on the region study, data available and also the work of Bergier et al. (2018), variables of interest were obtained from the National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL).

NOAA/ESRL data are obtained by gridded reanalysis data, which is a methodology from National Centre for Environmental Prediction (NCEP) that integrates various sources of data (aircraft, satellite, land observations, for example) into a single model through time. The model generation is automated and strictly monitored, with results being displayed after data gathering and collating at 2.5 by 2.5 degrees resolution for several variables (Kalnay et al., 1996; Wilberforce, 2007). According to Kalnay et al. (1996), gridded fields are classified in four subjective classes depending on the relative influence of the observational data and the model used: A, B, C and D. "A" indicates that the analysis is influenced almost exclusively by observational data, being the most reliable class; "B" indicates that exist observational data in the process, although results from analysis are included in the model; "C" includes no observational data, thus being derived solely from field models forced by data assimilation to remain close to the atmosphere and "D" are from fixed climatological values and is independent of the model. In this study we used only climatic variables which include observational data, from classes A and B. Table 2 summarizes the characteristics of the climatic variables considered in this study.

Table 2. Variables obtained from NOAA/ESRL website and its characteristics.

Climatic variable	**Description**	**Unit**	**Reliability class**
Air Temperature	Temperature of the air measured in a height of 1.5 meter above terrestrial surface.	Celsius (°C)	A
Precipitable Water	Total water vapour in atmosphere contained in a unitary section column between any two levels of surface, usually the top of atmosphere and terrestrial surface. It is the estimate of potential rain in determined region.	Kg/m^3	B
Meridional and Zonal Winds	The horizontal air movement relative to the terrestrial surface generated by atmospheric pressure gradients. Components of wind are direction, speed and force it exerts in determined object: - Meridional wind (v): positive values indicate winds coming from South; - Zonal wind (u): positive values indicate winds coming from West.	m/s	A

3.2 Methodology

This section shows how the time series were created. We assumed that the time series need to be equally spaced in relation to the observation collected time. However, since L1TP images can present cloud, not all images are suitable for analysis, then the time series obtained can be irregular. The problem, then, consists in selecting images in order to compose a regular sequence through time, from 1984 to 2017. The solution proposed was to select four images per year, so that each image represents one quarter of the year: the first one corresponds to information of January to March, the second from April to June, the third from July to September and the last from October to December. Yet, not all quarters were possible to be fulfilled due to the lack of image without clouds in some periods, so some kind of interpolation was necessary, which will be discussed in the next section. Time series were created and different aspects of the phenomenon of high macrophyte infestation were analyzed in three experiments, based on the time series constructed from the images:

Experiment I – Pixel QA assessment is used to inspect the quality of the portion of each temporal Landsat scene that included the study area.

Experiment II – NDVI/Landsat time series are used to investigate macrophyte dynamics throughout the reservoir and in specific sectors of the water body, defined based on the abundance and temporal persistence of their occurrence.

Experiment III – Time series of spectral indices extracted from images and climatic variables to analyze the influence of external factors on the dynamics of macrophytes in the reservoir.

3.2.1 Image pre-processing and selection of scenes for time series

Image pre-processing was done using the Geospatial Data Abstraction Library (GDAL) package. Each surface reflectance image (1-5, 7 for TM e ETM+, 1-7 for OLI) was pre-processed according to the flowchart shown in Figure 7. This was the initial steps carried for all experiments.

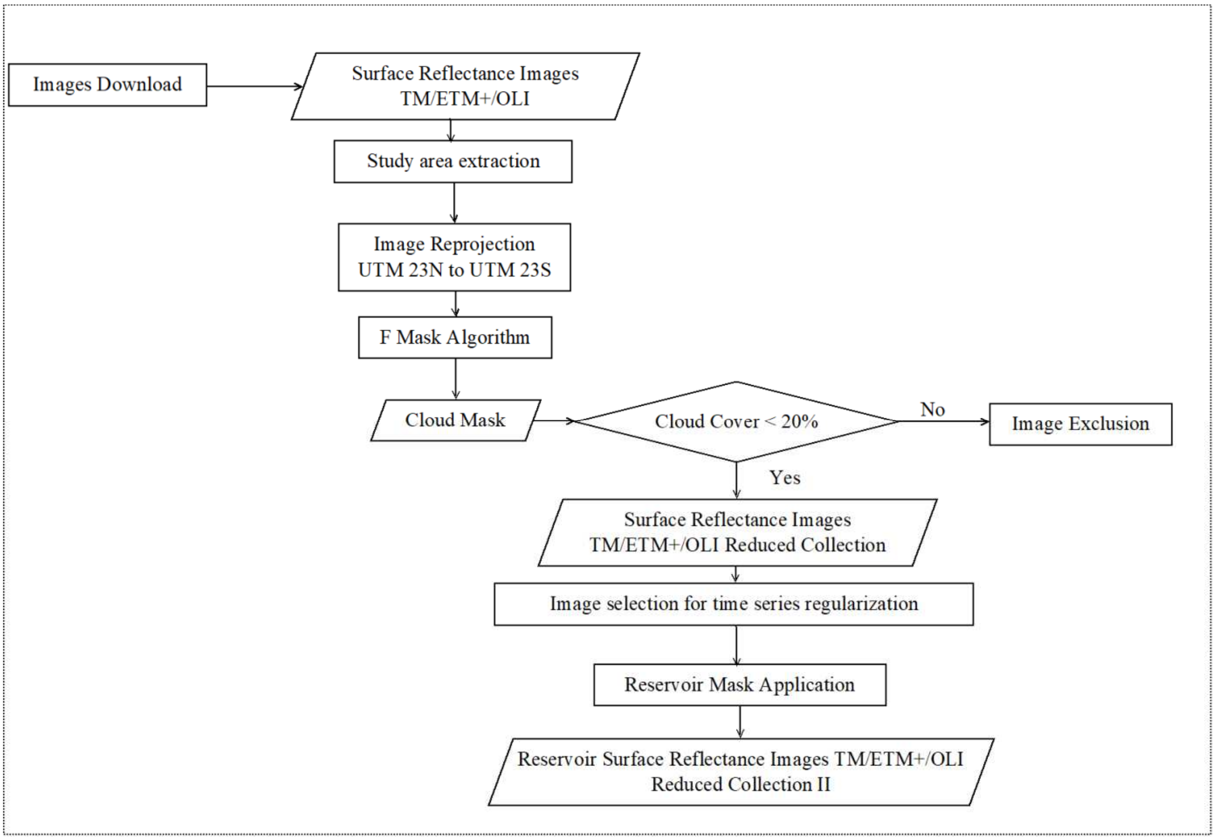


Figure 7. Methodological framework of general image pre-processing and selection of scenes for time series steps.

The downloaded images from 1984 to 2017 were first processed by clipping the area of interest (the reservoir) and re-projected to the Southern Hemisphere, UTM projection system, as indicated in Figure 7. Cloud cover was evaluated using the F-Mask algorithm (Zhu and Woodcock, 2012; Zhu et al., 2015), implemented in Python language, and scenes with of more than 20% cloud cover were removed, since it is considered a higher percentage (Li et al., 2018).

3.2.1.1 F-Mask algorithm application

The F-Mask algorithm was first published in Zhu and Woodcock (2012) and received improvements in Zhu et al. (2015). The algorithm is composed by a two-pass method for detecting clouds followed by an object-oriented method for detecting clouds shadows projected onto water and land surfaces (Lebreton et al., 2016).

By using NDVI and the top of atmosphere reflectance of NIR band, F-mask firstly divide all pixels into land and water pixels. Also, the algorithm assumes that the temperature from all clear-sky pixels in an image is approximately the same by using its brightness temperature to calculate cloud temperature probability. To define what clear-sky pixel is, the algorithm uses the first pass, which is a set of spectral tests that, if successfull, the pixel is considered as a clear-sky one. The second pass calculates normalized temperature probability, spectral variability and brighteness probabilities using thresholds generated automatically. For example, temperature probability uses the 17.5 and 82.5 percentiles of a sky-clear pixel's bright temperature. Then, a cloud probability mask is obtained and combined with the first pass clear-sky mask, and, from a dynamic threshold, cloud cover can be identified from cloud probability of all clear-sky pixels (Lebreton et al., 2016; Qiu et al., 2017).

Table 3 compares the number and volume of the past collection of images and the new collection, now with only the area of interest and images with 20% or less cloud information.

Table 3. Comparison of number of scenes and volume of image collection before and after using the F-mask algorithm.

	Before processing		After processing	
Satellite	Number of scenes	Volume (GB)	Number of scenes	Volume (GB)
TM/Landsat 5	315	560	183	2,92
ETM+/Landsat 7	294	608	178	3,16
OLI/Landsat 8	90	279	41	1,37

After applying the F-mask algorithm and obtaining the new reduced collection of images, we selected the ones used to compose the time series of images in a way that a regular time series is obtained.

In order to compose a regular time series from the cloud-free scenes, four images per year of the study period were selected, one for each quarter of the year. After May 2003, Landsat 5 TM images were selected over Landsat 7 ETM+, due to the failure of the ETM+ Scan Line Corrector (SLC), but it was necessary to use some scenes obtained by the latter sensor, corrected for the SLC-Off. Whenever no information was provided for pixels due to SLC-Off problem, estimation was obtained using a Kalman filter smoothing (Commandeur; Koopman, 2007) with values of the previous and next dates to interpolate the missing surface reflectance values.

3.2.1.2 Kalman filter application

The Kalman filter has the purpose of obtaining optimal values of a state at time point t only considering observations $\{Y(t): y_1, y_2, \dots, y_{t-1}\}$. A Kalman filtered state in time t, a_t, is given by the recursive scheme (considering a local model for simplicity):

$$a_{t+1} = a_t + K_t(y_t - a_t) \tag{7}$$

with K_t being the Kalman gain, which determines how much the prediction error at time t is allowed to influence the estimate of the state at time $t+1$: the larger the value, the larger the impact on the next filtered stage (Commandeur and Koopman, 2007). Considering a Kalman filter process at a time t, the current value of the filtered level based on all past observations before t is a_{t+1}, then, if y_t is missing in the time series, a_{t+1} contains all that could be learned from the past $\{y_1, y_2, \dots, y_{t-1}\}$ values, the best input for the filter will be $a_{t+1} = a_t$. However, if y_t is known, then the value can be fed into the Kalman filter to reduce discrepancies, i. e., $y_t - a_t$. Once the Kalman filter has been applied to the entire time series, inputing data for missing values of y_t are calculated by $\hat{y_t} = a_t$ (Durbin and Koopman, 2012).

To illustrate this process, Figure 8 shows a random time series with 100 points, where positions 33, 55 and 56 are missing, and thus Kalman filter was applied in those spots.

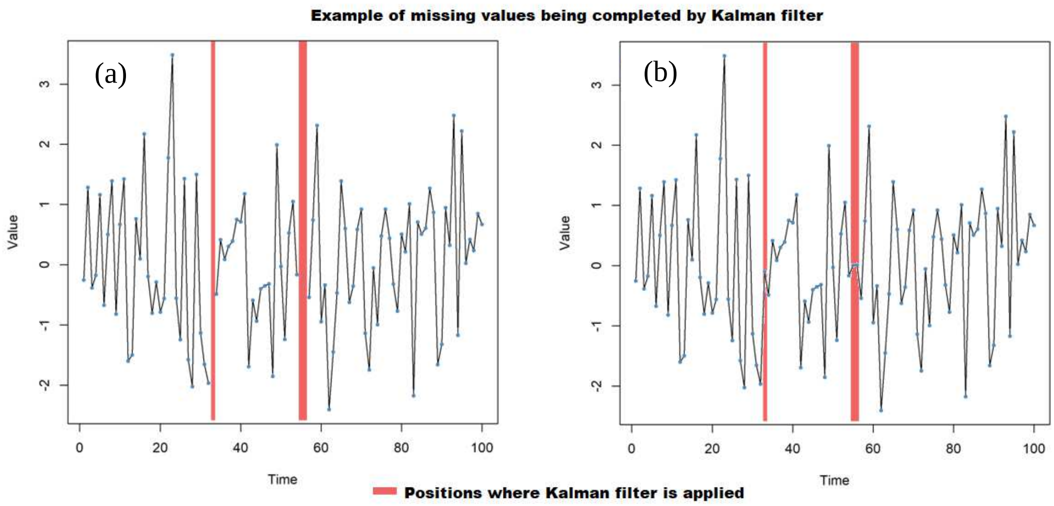


Figure 8. (a) Random time series with three gaps, identified by red bars. (b) Filled values in the time series by applying Kalman filter.

Table 4 presents a summary of the types of images used for each three-month period of the time series, wherein the empty boxes indicate the absence of a cloud-free image. For the time series regularization, data from the missing dates (empty boxes) were also estimated using the Kalman filter smoothing (Commandeur; Koopman, 2007) with values of the previous and next dates to interpolate the missing surface reflectance values.

Table 4. Landsat satellite sensors (TM (light grey), ETM+ (dark grey) and OLI (grey)) that collected each quarterly image. Q-1, Q-2, Q-3 and Q-4 represent each quarter of the year, where the empty boxes indicate no cloud-free image.

Q-4		TM	TM	TM	TM	TM	TM	TM	TM	TM		TM	TM		TM	TM	TM	ETM	ETM	TM	TM	TM	TM	TM	TM	TM	TM	ETM	ETM	OLI	OLI	OLI	OLI	OLI
Q-3	TM		TM	TM	TM	TM		TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	ETM	OLI	OLI	OLI	OLI	OLI
Q-2	TM	TM	TM	TM	TM	TM	TM	TM		TM	TM	TM	TM	TM	TM	TM	TM	TM	TM	ETM	TM	TM	TM	TM	TM	TM	TM	TM	ETM	OLI	OLI	OLI	OLI	OLI
Q-1		TM	TM	TM	TM		TM		TM	TM	TM		TM	TM	TM	TM	ETM	TM	TM	ETM	TM	TM	TM	TM	TM	TM	TM	TM	ETM	ETM	OLI	OLI	OLI	OLI
	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	1017

3.2.1.3 Reservoir mask application

A vector file of the contour of the reservoir, obtained through a RapidEye image from 03/08/2012, was used to create a mask of the water body, which was applied to each image of the time series. The option of RapidEye images for mask creating is based on their spatial resolution, which facilitates the contour drawing. Figure 9 illustrates the entire scene (a), results from the initial clip (b), a collection of clipped images (c) and the contour mask application, generating the analyzed image in the NDVI time series (d).

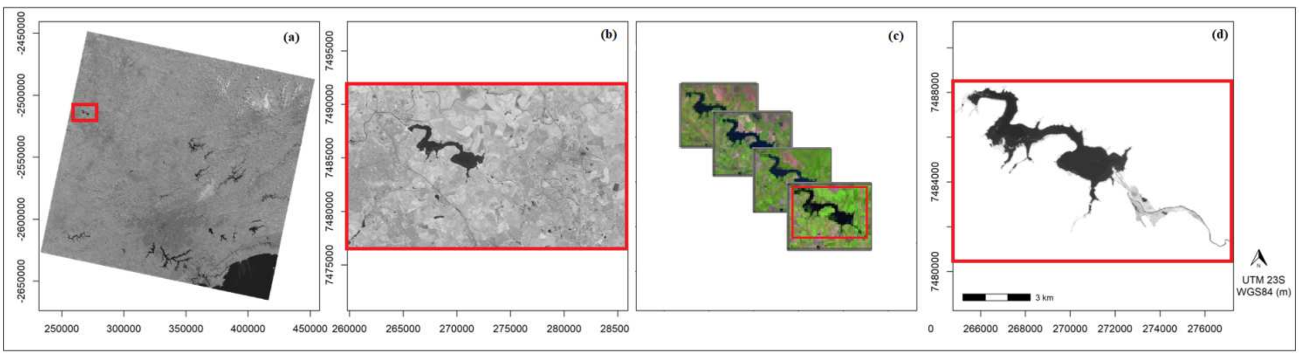


Figure 9. Delimitation of the study area from the path/row 219/76 Landsat scene; (a) Surface reflectance SWIR1 spectral band of the Landsat 8 OLI, (b) image resulting from the initial

image clipping; (c) clipped image representing different dates by Landsat color composite; (d) cropped image of the reservoir only area.

3.2.2 Experiment 1: assessing the quality of images in time series

The Pixel QA (Quality Assessment) Landsat product is also one of the possible downloadable bands from USGS Earth Explorer, which is a mask with each pixel containing a value originated by bits, which gives the type of possible obstruction present there: cloud, cloud shadow, snow, water or if it is clear. The value ranges are based on the radiometric resolution characteristics of the satellite sensors. While for Landsat 5-7 is expressed in 8 bits, Landsat 8 are in 12 bits (and scaled to 16), providing sixteen times the radiometric resolution of previous Landsat imagery (Yang et al., 2015, Li et al., 2018, Wulder et al., 2019). All QA information, however, applies to an entire Landsat image. A Landsat 8 image, for example, has a spatial resolution of 30m for multispectral bands and has a scene size of 170 km x 185 km (USGS, 2017). From this perspective, based on cloud, shadow or another obstructions presence, an overall quality score for the entire image may not reflect the quality for a particular small area. For example, cluster of obstructions still yields good parts of an image, while a scattering of obstructions might make the image worthless (USGS, 2018b).

This experiment was proposed to explore the potential of Landsat QA image-attributes obtained comparing results from entire images (as in Figure 8(a)) with results from the clipped images (as in Figure 8(c)). Furthermore, the assessment of quality discussed until now refers to a specific image and not for multiple images, thus, the idea was also to assess the quality in small area and for multiple images, in which information can be translated into a time series. The methodological framework of Experiment 1 is shown in Figure 10.

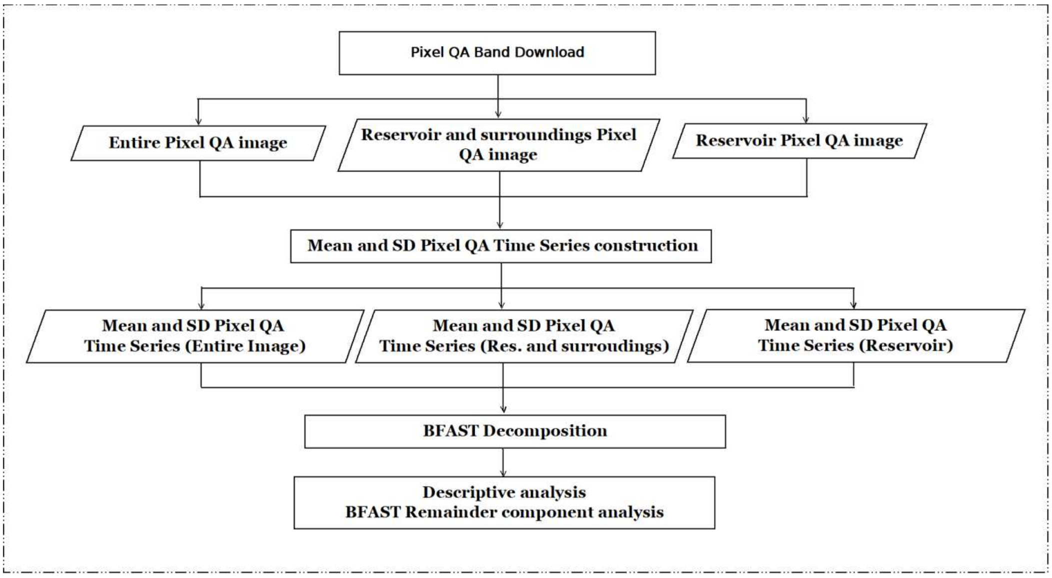


Figure 8. Methodological framework for Experiment 1.

3.2.2.1 Metadata usage

For this study, the QA-band information and interpretation using USGS' bits convention, which is different for Landsat 5-7 and 8, were used. Bits are set to 1 to denote surface state, cloud information or fill value and the position of the bit gives the proper interpretation, as given in Table 5 for both Landsat 5-7 and 8 quantizations (Dwyer et al., 2018).

Table 5. Bit interpretation for Landsat 5-7 and 8 for QA.

Bit	Bit Value	Cumulative Sum	Interpretation (L5/L7)	Interpretation (L8)
0	1	1	Fill	
1	2	3	Clear	
2	4	7	Water	
3	8	15	Cloud shadow	
4	16	31	Snow	
5	32	63	Cloud	
6, 7	64, 128	127, 255	Cloud confidence: 00=None, 01=Low, 10=Medium, 11=High	
8, 9	256, 512	511, 1023	Unused	Cirrus confidence 00=Not set, 01=Low, 10=Medium, 11=High
10	1024	2047	Unused	Terrain Occlusion
11, 12, 13, 14, 15	2048, 4096,8192, 16384, 32768	4095, 8191, 16383, 32767, 65535	Unused	Unused

Source: Adapted from USGS (2019a,c)

Using a combination of the values presented in Table 5, the values for the QA band itself are given in Table 6. As an example, the QA for cloud shadow, with low-confidence cloud for Landsat 5-7 is obtained setting bits 3 to 1, and bits 6 and 7 to 1 and 0, giving a value of 72.

Table 6. Pixel values for QA for Landsat 5-7 and 8.

Attribute	Pixel value (Digital Number) (L5/L7)	Pixel value (Digital Number) (L8)
Fill	1	1
Clear	66, 130	322, 386, 834, 898, 1346
Water	68, 132	324, 388, 836, 900, 1348
Cloud shadow	72, 136	328, 392, 840, 904, 1350
Snow/ice	80, 112, 144, 176	336, 368, 400, 432, 848, 880, 912, 944, 1352
Cloud	96, 112, 160, 176, 224	352, 368, 416, 432, 480, 864, 880, 928, 944, 992
Low confidence cloud	66, 68, 72, 80, 96, 112	322, 324, 328, 336, 352, 368, 834, 836, 840, 848, 864, 880
Medium confidence cloud	130, 132, 136, 144, 160, 176	386, 388, 392, 400, 416, 432, 898, 900, 904, 928, 944
High confidence cloud	224	480, 992
Low confidence cirrus	-	322, 324, 328, 336, 352, 368, 386, 388, 392, 400, 416, 432, 480
High confidence cirrus	-	834, 836, 840, 848, 864, 880, 898, 900, 904, 912, 928, 944, 992
Terrain occlusion	-	1346, 1348, 1350, 1352

Source: Adapted from USGS (2019a,b)

3.2.2.2 QA time series generation

To evaluate the influence of reservoir and surrounding areas size and the QA-bands in comparison with the full image provided by USGS site, in addition to the first mask produced in the pre-processing step, another vector file outlining the area surrounding the reservoir was used to create the appropriate mask to extract the area as shown in Figure 11(b). The region surrounding the reservoir boundaries was defined by the area of direct influence of the water body, according to Fonseca (2013) and represented previously in Figure 5.

In Figure 11 (a), an entire Landsat scene is shown highlighting the region where the reservoir is located, while in Figures 11 (b) and 11 (c) the areas of interest for the present study are shown: the surroundings near the reservoir and only the water body, respectively.

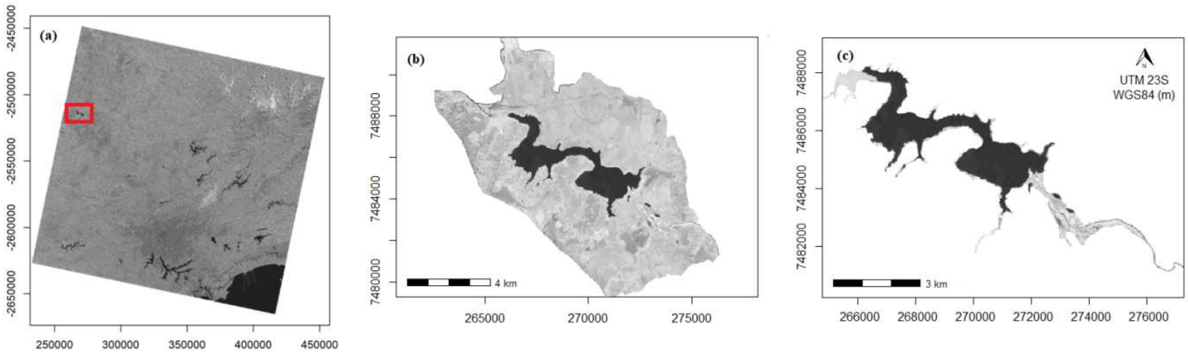


Figure 9. (a) Entire full TM/Landsat-5 image (NIR spectral band), highlighting the reservoir area. (b) Region surrounding the reservoir. (c) Reservoir region.

For each one of these images, the mean and standard deviation of QA information were calculated. To fill eventual missing values, Kalman filtering was applied to each time series in the missing spots to produce regularized quarterly 136-observation time series Pixel QA, as in Figure 9 example. Based on these values, a descriptive analysis was carried, as well as the BFAST decomposition of these series. Finally, assessing the variability by inspecting the remainder component of BFAST decomposition of the entire images, masks from Figures 11(a) and 11(b), some characteristics were found. The use of the remainder component was based on Verbesselt et al. (2010), underlying trend, when removed, improve the quality assessment as sensor drifts and calibration problems will not impact the signal received.

3.2.3 Experiment 2: trends and temporal changes based on NDVI time series

Considering the importance of evaluating the dynamics of occurrence of macrophytes in intensely anthropized small reservoirs, and seizing the availability of historical series of orbital multispectral images, this experiment proposes to explore the potential of Landsat image-attributes' time series to monitor the spatiotemporal behavior of aquatic macrophytes. Principal components analysis (PCA) and time series decomposition tools were exploited with the assumption that the combination of those techniques for analyzing large multi-temporal collection of images enables the analysis of trends and changes in the macrophytes occurrence in small reservoirs. Since NDVI is strongly correlated with biophysical vegetation parameters, it was used as a proxy to indicate periods of flowering or vegetation decline (de Jong et al., 2011).

The methodological framework of Experiment 2 is shown in Figure 12, where the Reservoir Surface Reflectance Images Reduced Collection II is resulted from the general pre-processing step discussed previously.

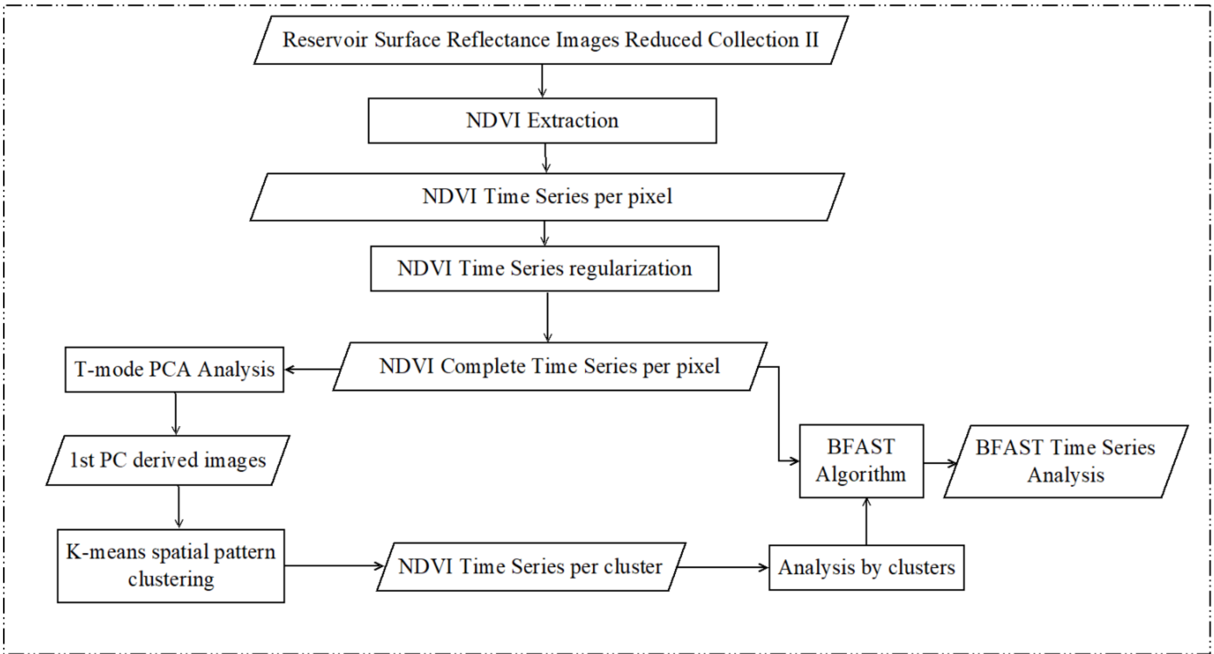


Figure 10. Methodological framework of Experiment 2.

3.2.3.1 Spatiotemporal analysis of NDVI

To produce a single image in which the occurrence of macrophytes in the Salto Grande reservoir is highlighted we used the NDVI spectral index, which has been applied in several local to global studies to detect and monitor changes in vegetation (see, e.g., Eckert et al., 2015). For long-term time series, periods indicated by growth and decrease of macrophytes are defined by the direction of the gradual change (regression slope) and are characterized as greening and browning, respectively (Jacquin et al., 2010; de Jong et al., 2013).

Each pixel of an NDVI image is characterized by its position (x, y) in the scene and an attribute representing the calculated NDVI values for that position. To describe the behavior of the NDVI and hence the occurrence of macrophytes in Salto Grande reservoir, temporal vectors are generated by extracting the pixel attributes at the same position along all images in time. In order to analyze the general spatiotemporal dispersion of this variable, choropletic maps representing the mean and variance of NDVI were plotted. However, it was necessary to develop a procedure to define specific sites to monitor the macrophytes temporal behavior. A transformation into principal component analysis (PCA) using the correlation matrix and the first principal component (PC1) was employed in the spectral decomposition of the multi-spectral and multi-temporal remote sensing images and reservoir sectors with similar temporal variability were defined using an unsupervised classification by the K-Means algorithm.

3.2.3.2 Principal Component Analysis

Principal component analysis is a statistical method for dimensionality reduction which transforms a set of p variables into a subset of q $(1 \leq q \leq p)$ principal components, with the first one explaining the largest portion of the variance in the original data, as well as they are being uncorrelated with each other (Johnson, Wichern, 2007; Kramer et al., 2019). Although PCA is usually performed on multivariate data through a matrix construction in which rows are the sample units and columns are the variables, it can be also used for single variables to identify dominant spatial patterns, which is the case of T-mode PCA, where the columns of the matrix are time points (Gallacher et al., 2016).

The basic building block of PCA is a matrix which contains the relationship among the variables (or time points) such as variance-covariance or correlation matrices. Given a square symmetric non-singular matrix $\boldsymbol{Q}$, the PCA transformation is built from the fact that $\boldsymbol{Q}$ is transformed into a diagonal matrix $\boldsymbol{\lambda}$, by pre- and pos-multiplying it by an orthonormal matrix of eigenvectors $\boldsymbol{E}$, i.e., $\boldsymbol{E}'\boldsymbol{Q}\boldsymbol{E} = \boldsymbol{\lambda}$.

The columns of $\boldsymbol{E}$ are the eigenvectors, which are the coefficients that transforms the original variables into new components in the manner of a weighted linear combination and the elements of $\boldsymbol{\lambda}$ are the eigenvalues, which determine the explanatory power of each of the components. The correlation between the original variables and a component is named loadings and the transformed value of any sample is named component score (Machado-Machado et al., 2011). In the case of T-mode, the results obtained by loadings are time series for each component that represents the correlation between each image and the component being diagrammed, with each variance explained by each component being the degree of significance of each spatial mode in the temporal domain. Finally, scores yielded by each principal component (PC) shows the spatial patterns which are strong or weak according to correlation given by the loading time series (Eastman and Fulk, 2003; Salles et al., 2001).

More recently, for example, Neeti and Eastman (2014) presented a PCA approach that compares spatial and temporal patterns among time series of multiple images. Adami et al. (2018) used PCA analysis in vegetation fraction images obtained by linear spectral mixture models in order to create a PCA-based color composite, and describe temporal seasonal changes in land cover and vegetation distribution. In this experiment, PCA is used to identify and to map the regions of similar temporal correlation in order to perform spatiotemporal analysis in the pixels aggregated in these regions, rather than single pixels. Considering the space-temporal dynamics of the occurrence of macrophytes in the reservoir, it was assumed that PC1 scores represent adequately the variability in the persistence and abundance of these plants in the water body.

Although PCA has been widely used in remote sensing, Jollife and Cadima (2016) pointed out an increased interest in PCA-related methods as a way to handle large data sets, including image processing and machine learning. In this experiment, PCA was applied as T-mode (Neeti and Eastman, 2014) to define subsets of pixels that have similar spatial variability over time, since the focus of this mode is to identify dominant recurring spatial patterns over time. Note that PCA was applied by considering each time series image as variable and its pixel attribute (NDVI) as observation. The loadings of the PC1 defined a temporal profile, whose outliers represent moments in the time when an unusual behavior of the variable or a problem in the image occurred.

3.2.3.3 K-Means Clustering Algorithm

K-means clustering is an unsupervised machine learning partitioning-based grouping technique, which is based on the iterative relocation of data points among clusters whose goal is to produce groups of cases or variables with a high degree of similarity within each group and a low degree of similiratity between the groups (Morissette and Chartier, 2013). The algorithm defines a prototype in terms of a centroid, usually the mean of objects in continuous n-dimensional space (Tan et al., 2019).

The basic algorithm starts by a user-specified parameter k, which is the initial number of centroids, i.e., the number of clusters of interest. There are also techniques in choosing a proper k value, as can be seen in Pham et al. (2004), for example. Then, each point is assigned to the closest centroid and each collection of points that is close to the centroid defines a new cluster, which is updated based on the points assigned to this cluster. The process is repeated until no point changes clusters, i. e., until the centroids remain the same.

Mathematically, to partition n objects $x_i, i \in \{1, 2, \dots, n\}$ into k clusters in which each object belongs to a cluster with the closest mean and considering $c_j, j \in \{1, 2, \dots, k\}$ is the centroid for cluster j, the objective is to minimize the total within cluster variance by the squared error function

$$\sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (8)$$

in which $\left\| x_i^{(j)} - c_j \right\|^2$ is the distance function. The Euclidean distance is usually used to assign data points to the closest centroid (Tan et al., 2019).

The scores of the PC1 are submitted to a classification by the K-means algorithm, allowing for clustering pixels that present similar scores and define the regions of the reservoir for the spatio-temporal analysis. For each cluster, the mean NDVI of the included pixels was calculated and the mean NDVI time series constructed.

3.2.3.4 Time series analysis using BFAST

Due to its definition as a temporal vector, the NDVI can be analyzed as a time series, in which $T = \{t : t_1 < t < t_2\} = \{1984, \dots, 2017\}$ is the set of values of the variable analyzed in time and {Y(t):t∈T} is the value assumed by the variable NDVI in a specific time. The BFAST algorithm allows capturing gradual and abrupt changes (breakpoints in the time series) in long-term series by decomposing the series in three components: trend, seasonality and remainder. Although several other approaches have been proposed to study satellite images, for example, by applying wavelet decomposition (Martínez and Gilabert, 2009) or only PCA (Rieser et al., 2010), these methods involve some type of transformation to isolate dominant aspects with the challenge of labeling and identifying the change components. (Verbesselt et al., 2010). BFAST algorithm was chosen for this study because it was primarily developed and tested by simulating NDVI time series with varying quantities of seasonality and noises, in addition to different abrupt changes. The component of seasonality is defined as a sequence of harmonic terms, while continuous trend behavior is evaluated and straight segments are adjusted piecewise. Thus, trend is composed of segments representing gradual changes, separated from each other by abrupt changes (Jamali et al., 2015). This piecewise modeling of the trend is often used to approximate complex phenomena and extract basic data characteristics (Zeileis et al., 2003).

In time series of variables extracted from satellite images, BFAST has mainly been used in studies on vegetation cover changes. Darmawan and Sofan (2012) applied the algorithm to detect changes in tropical rainforest, where they compared the performance of two vegetation indices, NDVI and Enhanced Vegetation Index (EVI). Watts and Laffan (2014) used BFAST approach to detect vegetation response in semi-arid regions of Australia by studying long term trends and magnitudes. Tian et al. (2015) evaluated the temporal consistency of multi-sensor NDVI time series using BFAST by analyzing the occurrence between breaks in time series and sensor shifts. Permatasari et al. (2016) applied it in agricultural land use changes in Indonesia using EVI and K-means clustering patterns to detect temporal and spatial patterns.

The analysis of trend and seasonality in this experiment, as well as breaks in the time series, considered regions of homogeneous temporal variability, spatially aggregated by means of a K-means algorithm. All of those steps, besides the map creations, were implemented in software for statistical computing (R Development Core Team, 2016) using *Rgdal, raster, bfast* and *psych* packages.

3.2.4 Experiment 3: influence of exogenous variables on the spatiotemporal behavior of macrophytes

Aside from anthropogenic action, when aquatic vegetation is grown in reservoirs, environmental factors can interfere in the macrophytes growth behavior as well. As pointed by Lacoul and Freedman (2006) and Mäkela et al. (2004), many studies confirmed the influence of interacting environmental factors on the distribution of aquatic plants in lentic and lotic waters: exogenous environmental variables in macrophytes context were explored by Jacquin et al. (2010), Fusilli et al. (2013) and Eckert et al. (2015). Then, it is essential to monitor the distribution and variations of aquatic macrophytes (Luo et al., 2016) and to understand and quantify the environmental factors that influence their distribution patterns (Dar, Pandit and Ganai, 2014). In contrast, as reported by Zhang et al. (2016), small reservoirs still receive little attention, while large ones are regularly monitored. Moreover, Chandler et al. (2018) emphasized that remote sensing is mostly used for monitoring lakes macrophytes' conditions, with rivers receiving almost no research attention.

Thus, since on one hand the importance of evaluating the macrophytes distribution is considered, while on the other hand there is a lack of information in small reservoirs located in anthrophized areas, this experiment proposes to extend Experiment 2 with the addition of exogenous climatic variables. It is hypothesized in this experiment that the dispersion of macrophytes banks is influenced by climatic conditions such as temperature, precipitation, wind direction, beyond the surrounding land use, and proposed a methodology to investigate similar phenomena in small reservoir areas.

The methodological framework of Experiment 3 is shown in Figure 13, where the pre-processing step is common from Experiment 2.

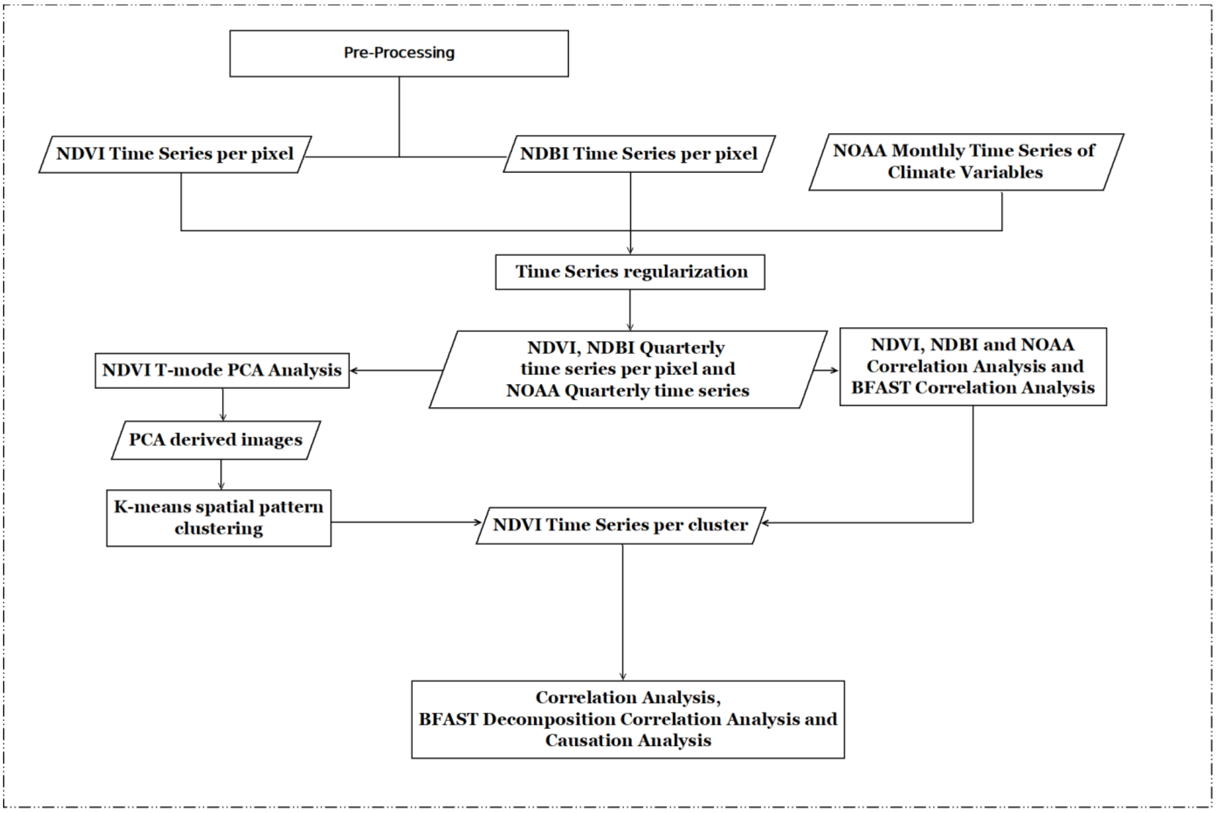


Figure 11. Methodological framework of Experiment 3.

3.2.4.1 Exogenous climatic variables

To evaluate the influence of exogenous variables on the dispersion of temporal changes in the macrophyte occurrence in the reservoir, a time series analysis approach based on attributes extracted from Landsat images and climatic variables was designed to investigate two scenarios. Initially, the influence of anthropogenic occupation around the reservoir was considered, comparing the NDVI and the NDBI time series, associated with the degree of imperviousness. The influence of weather conditions was evaluated in a scenario in which the NDVI time series from the reservoir was compared with the climatic oscillation time profiles derived from the NOAA historical records of the main climatic variables.

3.2.4.2 Defining horizontal sectors in the reservoir

To evaluate the influence of anthropogenic occupation in the surroundings of the reservoir on the macrophyte dispersion in the water body and the effect of climatic variables on the behavior of these aquatic plants, two masks previously created were used to extract the appropriate areas in the images: vector files outlining the area surrounding the water body (Figure 14(b)) defined according to Fonseca (2012) and shown in Figure 5, and the reservoir itself (Figure 14(c)), based on the RapidEye image from 03/08/2012 and already indicated in Figure 8 (d).

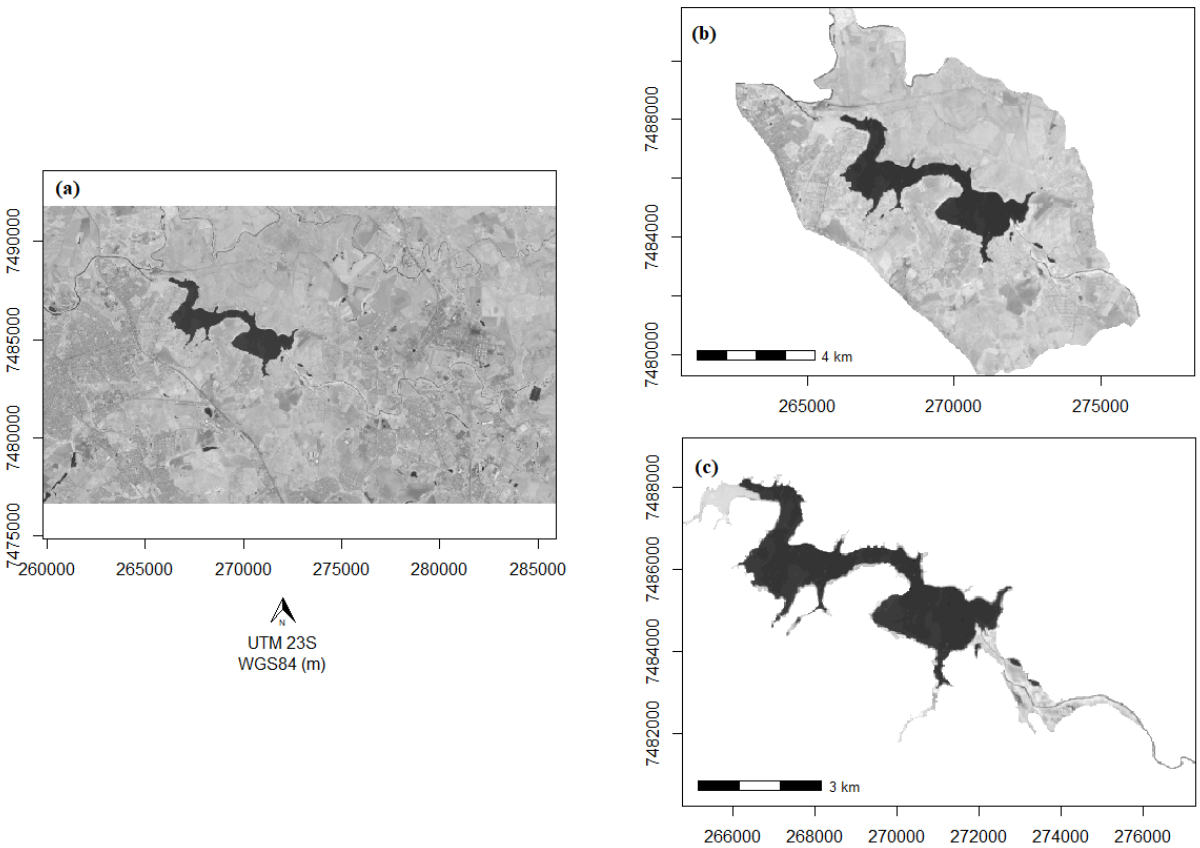


Figure 12. Area of interest cropped from an entire TM/Landsat-5 scene, NIR spectral band (a), region surrounding the water body (b), and reservoir region (c).

The purpose of the application of mask in Figure 14(c) was to analyze the reservoir itself without urban influence around the reservoir (which is done with mask from Figure 14(b)), decreasing the ambiguity of macrophyte occurrence and other vegetations like agricultural culture near the reservoir. Using the procedures of Experiment 2, PC transformation was applied in maps for mask in 14(c) and PC1 was used to identify and map the regions of similar temporal correlation. The scores of this first component were submitted to a classification by the K-means algorithm, allowing the creation of pixel clusters which presented similar scores and, thus, the definition of areas in the reservoir for further spatiotemporal analysis. Twelve clusters of interest (which will be named as “sectors”) were selected for inspection. For each area, the mean NDVI values were calculated and the mean NDVI time series were built. NDVI for the entire reservoir area was also calculated (referred as global NDVI). Finally, groups of interest (named “areas”) were created as “near dam”, “reservoir body” and “near the river” sectors.

3.2.4.3 Time series analysis and Granger causalities

Time series analysis was executed in two stages: application of the BFAST decomposition algorithm to the time series, and testing if meteorological variables helped the forecast of NDVI by using the Granger-causality concept.

The Granger causality technique (Granger, 1969) is used to study if a time series of a specific variable is useful in forecasting another variable. This technique has been used in addressing causality problems in climatic systems (Attanasio et al., 2013). If X_t and Y_t are time series, Granger (1969) uses the expression “X_t Granger-causes Y_t” to express the fact that the knowledge of past values of X_t reduces the variance of the errors in forecasting the values of Y_t beyond the variance of the errors which would be made from the knowledge of past Y_t alone (Norrulashikin et al., 2016). In mathematical notation, two models are considered, namely restricted (Equation 9) and unrestricted (Equation 10) models, given by:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \epsilon_t \tag{9}$$

$$\begin{aligned} Y_t = \beta_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} \\ + \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} + \gamma_t \end{aligned} \tag{10}$$

with α_0 and β_0 being constants terms, α_i and β_i for $i = 1, \ldots, p$ the unknown parameters, ϵ_t and γ_t white noise errors and p the maximum lag allowed for both models. To test if the restricted estimates are statistically different from the unrestricted estimates, the F-test $F = \frac{(SSE_r - SSE_u)/p}{SSE_u/(n-k)}$ is used, with SSE_r and SSE_u being the sum of squared errors of restricted and unrestricted models, respectively, k the number of coefficients in the unrestricted model and n is the sample size. If $F < F_{H,n-k}$, $F_{H,n-k}$ being the tabulated value of a F-distribution with parameters p and $n - k$, it is concluded that X does not Granger-cause Y.

For this study models were investigated by adding lags until significance is observed at 0.05 level, which are shown in equations (11) to (14).

$$NDVI_t = \alpha_0 + \sum_{i=1}^{4} \alpha_i NDVI_{t-i} + \sum_{i=1}^{4} \beta_i Air\ temp_{t-i} + \eta_t \tag{11}$$

$$NDVI_t = \alpha_0 + \sum_{i=1}^{4} \alpha_i NDVI_{t-i} + \sum_{i=1}^{4} \beta_i Precip\ water_{t-i} + \epsilon_t \tag{12}$$

$$NDVI_t = \alpha_0 + \alpha_1 NDVI_{t-1} + \beta_1 Merwin_{t-1} + \Gamma_t \tag{13}$$

$$\begin{aligned} NDVI_t = \alpha_0 &+ \sum_{i=1}^{2} \alpha_i NDVI_{t-i} + \sum_{i=1}^{2} \beta_i Air\ temp_{t-i} \\ &+ \sum_{i=1}^{2} \gamma_i Precip\ water_{t-i} + \sum_{i=1}^{2} \delta_i Merwin_{t-i} + \zeta_{\mathrm{t}} \end{aligned} \tag{14}$$

with $NDVI$ being global NDVI values, $Air\ temp.$ the values for air temperature, $Precip\ water$ values for precipitable water and $Merwin$ the meridional wind values, all for $t = 1984, \ldots, 2017$, α, β, γ and δ the unknown parameters estimated and η, ϵ, Γ and ζ white noise errors.

4 RESULTS AND DISCUSSION

4.1 Assessing the quality of images in time series to analyze the spatiotemporal dynamics of aquatic macrophytes in small eutrophic reservoirs

For each mask presented in Figure 11(b) and (c), as well as for the entire image (Figure 13, left), the mean of all QA pixels was calculated. Table 7 contains the legend for Tables 8(a)-(c). The numbers depicted in Tables 8(a)-(c) represent the mean of all QA pixels for each image (where no image was available, values were interpolated using Kalman filter) and the colours represents the closest possible interpretation of that number (USGS, 2019a). As mentioned previously, due to the distinct quantization used by Landsat 5-7 and 8, higher values of digital numbers are observed for Landsat 8 and thus, additional classifications were added to the Table 7.

Table 7. Classification of QA pixels values for the QA mean pixels value for studied images. Labels refer to the reference satellite for the color: L5/7 is Landsat 5 and 7, and L8 is Landsat 8.

L5/7	Clear or water, low-confidence cloud
L8	Clear or water; low-confidence cloud; low-confidence cirrus
L5/7	Cloud shadow; low-confidence cloud
L8	Cloud shadow; low-confidence cloud; low-confidence cirrus
L5/7	Cloud, medium-confidence cloud
L8	Cloud, low-confidence cloud; low-confidence cirrus
L5/7	Cloud shadow, medium-confidence cloud
L8	Cloud, medium-confidence cloud; Cirrus, low-confidence cirrus
L5/7	Cloud, high-confidence cloud
L8	Cloud, high-confidence cloud; Cirrus, low-confidence cirrus

SOURCE: Adapted from USGS (2019b,c)

Table 8. Series of (a) entire Landsat images mean QA pixels, (b) Reservoir plus surroundings images mean QA pixels, and (c) Reservoir images mean QA pixels.

	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
a																	
Q-1	160	69	71	73	85	77	77	106	79	174	70	104	76	82	76	68	84
Q-2	171	175	94	87	93	79	90	148	109	101	96	144	71	127	154	92	90
Q-3	105	96	73	79	82	69	97	73	85	140	70	128	102	150	107	88	145
Q-4	123	101	101	77	98	87	114	171	90	171	107	114	85	121	105	99	128
b																	
Q-1	64	66	66	66	66	59	66	59	66	66	66	59	66	66	66	66	66
Q-2	66	224	66	66	66	66	68	66	70	66	66	66	66	66	66	66	66
Q-3	66	66	66	66	66	66	68	66	66	171	66	66	66	66	66	66	66
Q-4	86	86	66	66	66	66	66	66	67	66	67	67	66	69	66	66	66
c																	
Q-1	65	68	68	68	68	60	67	60	68	67	67	60	67	67	67	67	67
Q-2	68	224	67	68	67	68	72	68	71	67	67	67	67	67	67	67	67
Q-3	68	68	68	68	68	68	70	68	67	186	67	67	67	66	67	67	67
Q-4	88	88	68	68	67	67	67	68	71	67	68	67	67	70	67	67	67

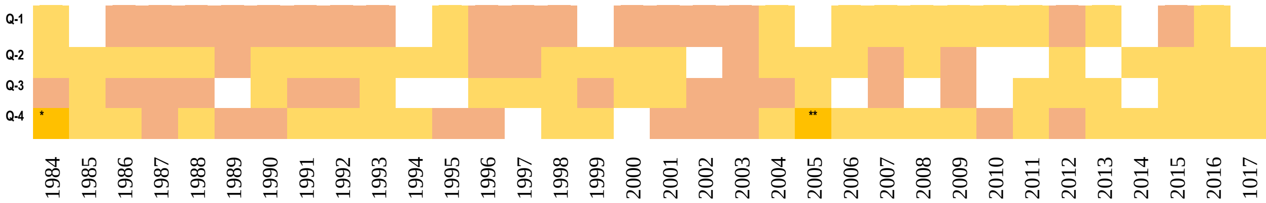
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	1017
a																	
Q-1	77	74	71	115	135	97	97	114	106	102	116	82	213	323	330	384	380
Q-2	94	67	81	95	93	104	85	100	89	67	69	156	326	538	372	343	337
Q-3	92	71	86	88	108	66	79	66	79	66	95	100	330	325	347	352	329
Q-4	71	75	76	91	78	102	107	93	90	72	145	77	362	412	503	398	337
b																	
Q-1	66	67	66	66	66	66	68	66	66	66	66	66	78	322	322	322	322
Q-2	66	67	66	66	66	66	66	66	66	66	67	66	322	322	322	322	322
Q-3	66	66	66	66	66	66	66	66	67	66	83	66	322	322	322	322	322
Q-4	66	66	66	66	87	66	78	74	73	66	93	66	322	322	322	322	322
c																	
Q-1	67	67	67	67	67	67	73	67	67	67	67	67	82	323	323	323	323
Q-2	67	67	67	67	67	67	67	67	67	67	68	67	323	323	323	323	323
Q-3	67	67	67	67	67	67	67	67	67	67	87	67	323	323	323	323	323
Q-4	67	67	67	67	93	67	105	75	80	67	90	66	323	323	323	323	323

When compared to Tables 8(a), 8(b) and 8(c) it is possible to note the larger variation in Table 8(a) than in the others, as several mean values showed cloud cover or shadows. From Tables 8(b) and 8(c) it is observed that the mean QA values are similar for almost the entire dataset, despite the fact that a larger portion of the study area is considered in the second table, that is, the reservoir together with its surroundings (Figure 14(a)). For the most part, a clear pixel is identified with low confidence cloud (the probability of cloud is low), meaning that, although the values from Table 8(c) differ from the first two, the reservoir area (and its surroundings) are not affected by eventual cloud or shadow problems which can be present in the entire image.

From Table 8(a) and 8(c), eighteen values out of the 136 (13.23%) remained in the "clear/water" pixel category after the image was clipped. For example, Q1-1985 had a value of 69 in Table 8(a) and in 8(c), the value is 68. Q3-2014 is the value 325, while in Table 8(c) it is 323. Only one value (0.73%) that was categorized "clear/water" pixel in Table 8(a) became "cloud shadow" in Table 8(c), although with low confidence, which was in Q4-1984. A value in Q4-2005 remained as "cloud shadow" after the image change as well. Similar results were obtained by comparing Table 8(a) with 8(b).

In most of cases, though, categories like "cloud" or "cloud shadow" appearing in Table 8(a) became "clear/water" pixels in Tables 8(b) and 8(c). For "cloud shadow" 44 values (32.35%) became "clear/water" pixels, and 70 values (51.47%) went down from a higher confidence to a low confidence for clouds or cloud shadow. Table 9 shows which images changed its confidence classification or maintained it by inspecting values of Table 8(a) and 8(b). Inspecting Table 8(a) and 8(c) produced similar results.

Table 9. QA values which changed its confidence status or maintained it when comparing Tables 8(a) and 8(b).



	From high/medium confidence to clear pixel.
	Maintained its classification of clear pixel.
	From low confidence to clear pixel.
	*From clear to cloud shadow; **Maintained as cloud shadow.

When comparing to the original image time series, 20 images out of 136 had the classification of a "clear/water pixel" (14.70%) and 55 had the classification of cloud shadow with low confidence (40.44%), meaning that 55.14% of the entire images have a clear condition or low probability of clouds. However, all images selected were considered with 20% cloud cover or less with F-mask algorithm. Then, considering the mean value of QA pixel for the entire images need caution when assessed, as information from F-mask algorithm can improve the clear percentage. Now, considering the clipped images for reservoir and its surroundings, and only reservoir, both are 98.53% composed of "clear/water" or low confidence cloud shadows, with the reservoir only image containing more of "cloud shadow" classification.

Figure 15 shows the time series for Mean QA pixels, as well as Standard Deviation (SD) and respectively BFAST decompositions for (a) entire Landsat image, (b) reservoir and its surroundings, and (c) only reservoir. Note that the increases in values after Q1-2013 in the original mean time series are due to the fact of changing satellites for OLI imagery.

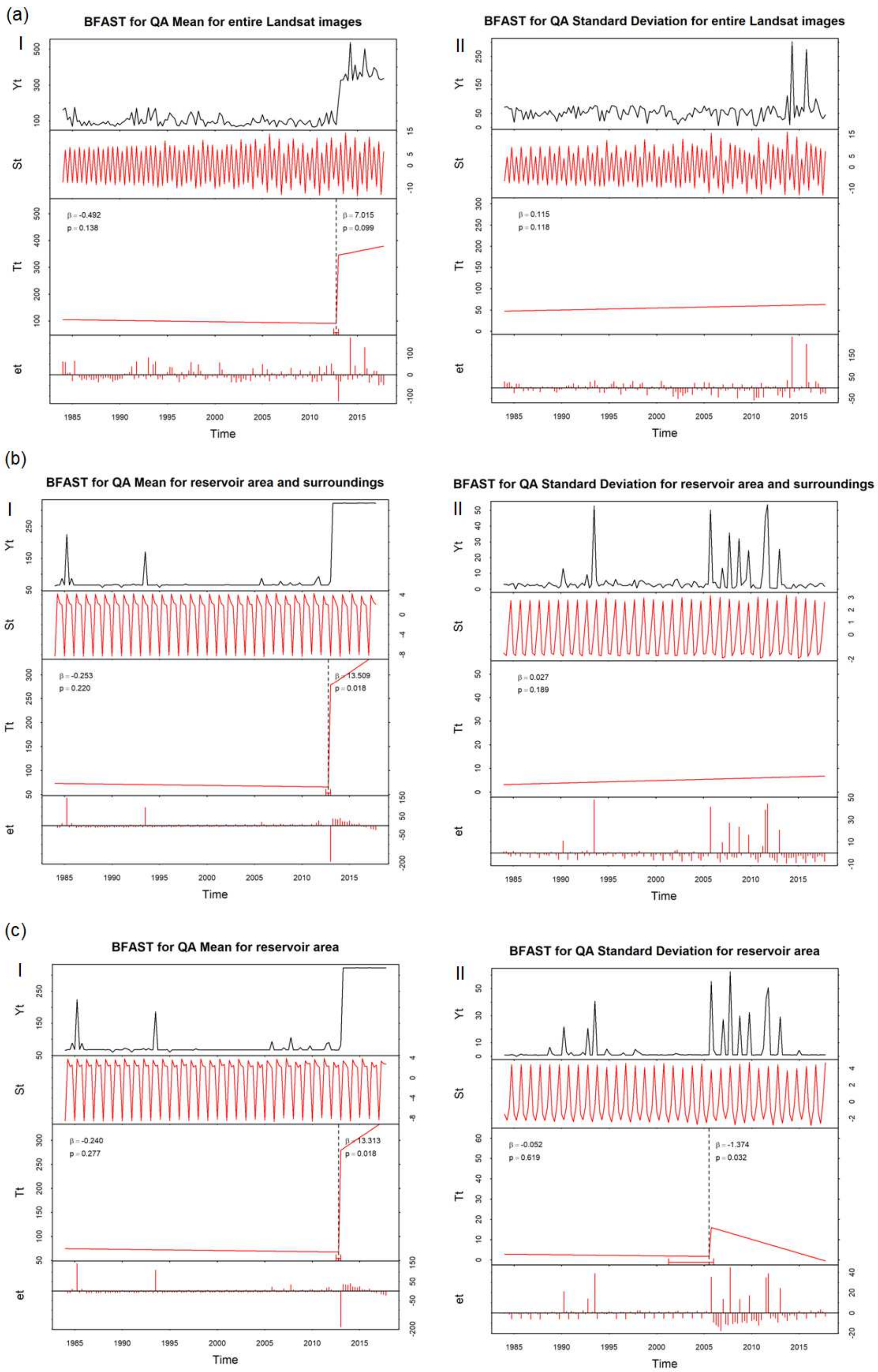
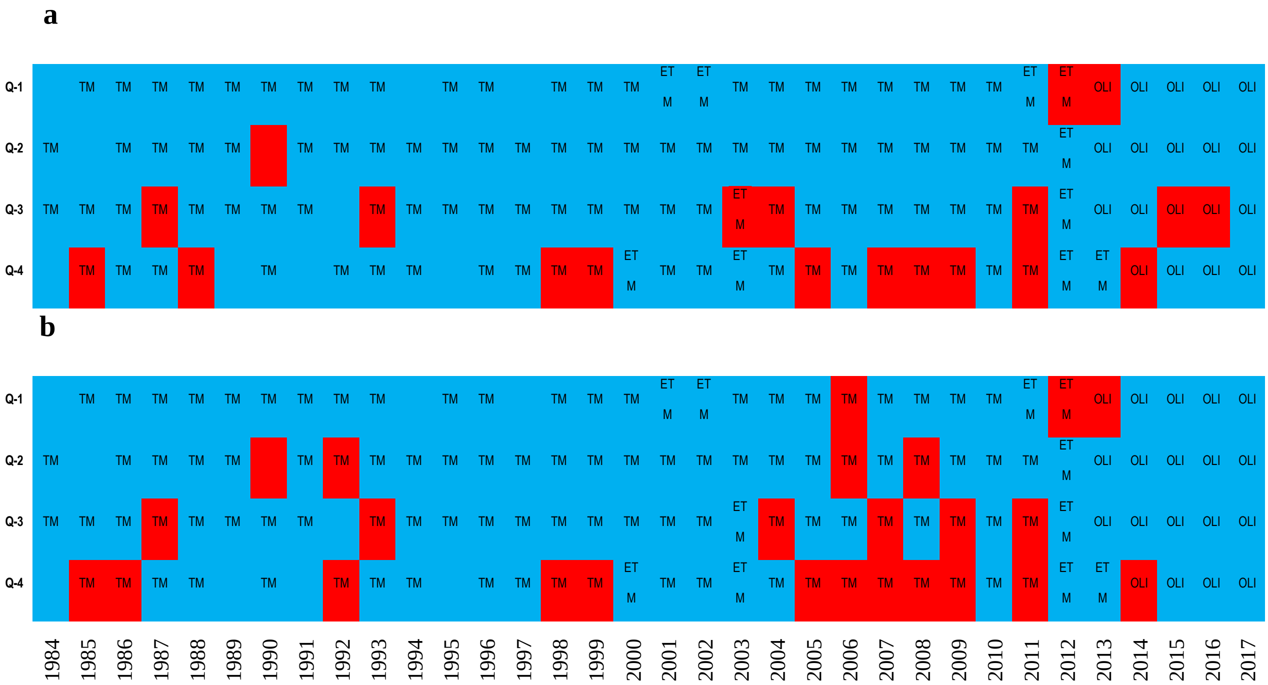


Figure 13. BFAST applied to (a) entire Landsat, (b) reservoir and surroundings, (c) reservoir time series of (I) QA mean and (II) QA standard deviation. For each graph, Y_t represents the original time series, S_t is the seasonality component, T_t is the general trend (where ρ is the slope and p is p-value).

After decomposing the QA mean and QA standard deviation time series through BFAST in Figure 15, the number of images of reservoir plus its surroundings and images of only reservoir which remainder component of standard deviation obtained by BFAST decomposition were lower or equal than the remainder component of the entire images was verified. Table 10 summarize the results indicated by colours: blue if the SD is lower or equal, red if the SD is higher.

Table 10. (a) Reservoir plus its surrounding images and (b) Reservoir images which had lower remainder component of standard deviation (in blue), or higher (red) in comparison to the entire image (TM – Thematic Mapper, ETM – Enhanced Thematic Mapper Plus, OLI – Operation Land Imager).



From Table 10(a), the standard deviation errors, 85.3% (116 images) of the surroundings and reservoir images had a smaller value of SD than the entire image. For (b), the standard deviation errors, 81.61% (111 images) of the reservoir images had a smaller value of SD than the entire image. It was also observed that the majority of images which SD is higher than the original image were found in Q4 period and concentrated in the period of 2004-2011, period where ETM+ images were already available.

As an example, Figure 16 shows the full image, reservoir and surroundings image, and only reservoir image for the date for 2005 Q4. From Tables 8(a)-(c), all had low-confidence cloud shadow, and from Tables 10, both sub regions (reservoir and surroundings, and only reservoir) showed more residuals than the original image, although the same category of image was obtained: “cloud shadow”.

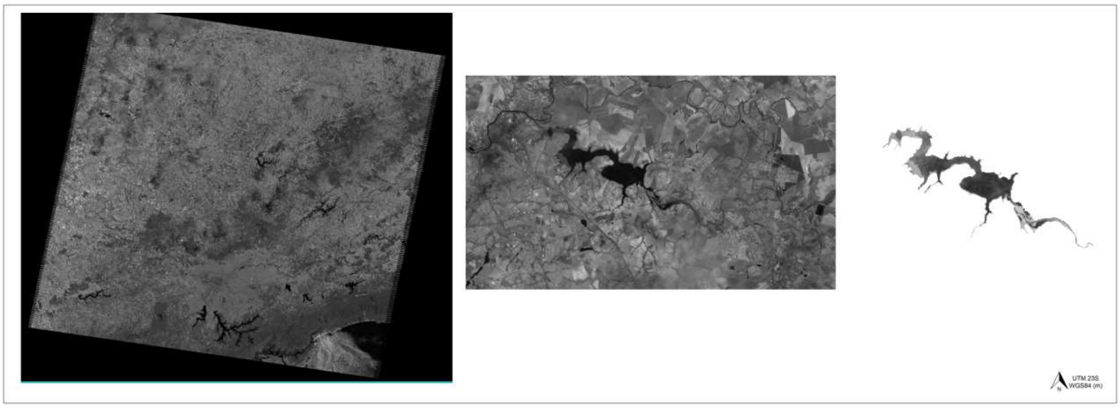


Figure 14. Full Landsat 5 image, reservoir and surroundings image, and only reservoir image for 20005Q4.

From Tables 10(a)-(b), the quality information of clipped areas could be improved based on low values of standard deviation and the changing of cloud confidence, for example, by “flagging” which images contain which one of information. This showed that using BFAST decomposition in the Pixel QA quality band is a possibility to assess the quality of parts of a Landsat satellite image and can also possible be an interesting new use for the algorithm, since it is specified by Verbesselt et al. (2010b) that BFAST can detect various types of changes, but it is important to understand how these changes will be detected.

4.2 Trends and temporal changes in the macrophytes abundance in an eutrophic tropical reservoir by long-term Landsat time series

4.2.1 Macrophytes persistence based on NDVI spatiotemporal distribution

Figures 17(a) and 17(b) show the spatial distribution of the mean NDVI and the variance over time indicating the regions of the reservoir with persistent and variable occurrences of macrophytes.

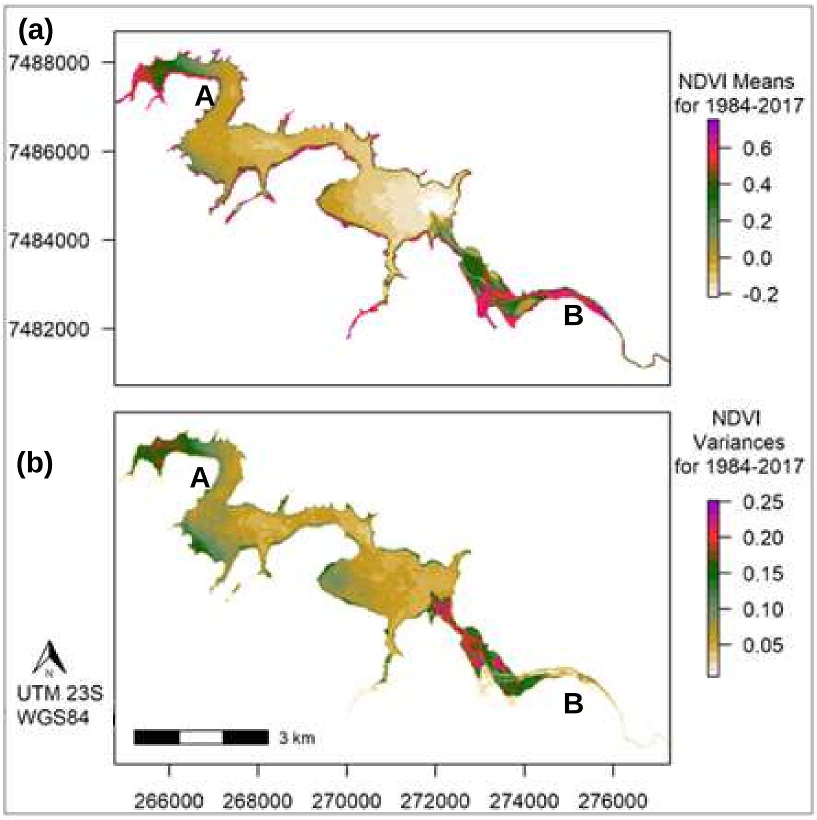


Figure 15. Spatiotemporal distribution of (a) the mean NDVI and (b) corresponding variance in the Salto Grande reservoir in the period 1984-2017. Area A represents the region near the dam and Area B represents the region near the river.

Three main conditions of macrophyte occurrence in the Salto Grande reservoir can be characterized: (i) the abundance and systematic occurrence of plants (NDVI higher than 0.4) near dam (area A) and near Paulínia city (area B); (ii) irregular occurrence of macrophytes (mean NDVI close to 0.2), and (iii), absence of plants on the eastern side of the lake in B. The highest NDVI values were observed in in areas A and B, indicating constant and abundant presence of macrophytes over time, close to the urban agglomerates (cities of Americana and Paulínia, respectively) and are the outcome of high degree of pollution diffuse. For these same areas (A and B), Figure 17(b) shows that in the places where inputs of the tributaries occur, the NDVI variability is low. This means that these are locations with large occurrence of macrophytes. The largest NDVI variances are found in region B, suggesting that there is a seasonal displacement of macrophytes, influenced by internal and external factors to the water system. In the recesses that occur in the northwest side of the reservoir, variances between 0.1 and 0.15 were observed, indicating the periodic presence of macrophytes, probably associated with the variation following the direction of the wind (as pointed by e.g. Leite, 1998; Martins et al., 2011).

4.2.2 Time series outliers identification in the PC1 loadings

By applying PCA in the NDVI dataset, the four variances obtained for the first five principal components are listed in Table 11.

Table 11. Proportion of variance and proportion explained by the first five PCs.

	PC1	PC2	PC3	PC4	PC5
Proportion Variance (%)	61	8	4	3	2
Cumulative Variance (%)	61	68	72	75	77
Proportion Explained (%)	79	10	5	4	2
Cumulative Proportion (%)	79	89	94	98	100

The proportion of variance retained in PC1 was 61% indicating that NDVI values fluctuated over time, confirming the variability in spatial distribution and abundance of aquatic macrophytes in the reservoir. Considering PC2, PC3, PC4 and PC5, they had low explanation power towards the variance as they summarize only 17% of total proportion variance. Due to the proportions retained from the first PC and the lower proportions from the others, only the first component was used for the rest of this study, in order to sectorize areas of the reserve as to the temporal correlation in the occurrence of macrophytes.

The temporal profile of the first principal component loadings is shown in Figure 18(a), which highlights the points in which the correlation between PC1 and a specific date was less than 0.5, considered outliers. The NDVI images corresponding to the dates coincident with these outliers are shown in Figure 18(b), allowing the identification of atypical behaviors in the image or in the dispersion of macrophytes in the reservoir. From the PC1 loadings time series, it is possible to verify that most temporal NDVI images presented positive correlation with PC1, some of them larger than 0.7. This means that the spatial pattern observed in Figure 17(a) is similar in several time periods of the study, except in the case of outliers, when the correlation decreased considerably.

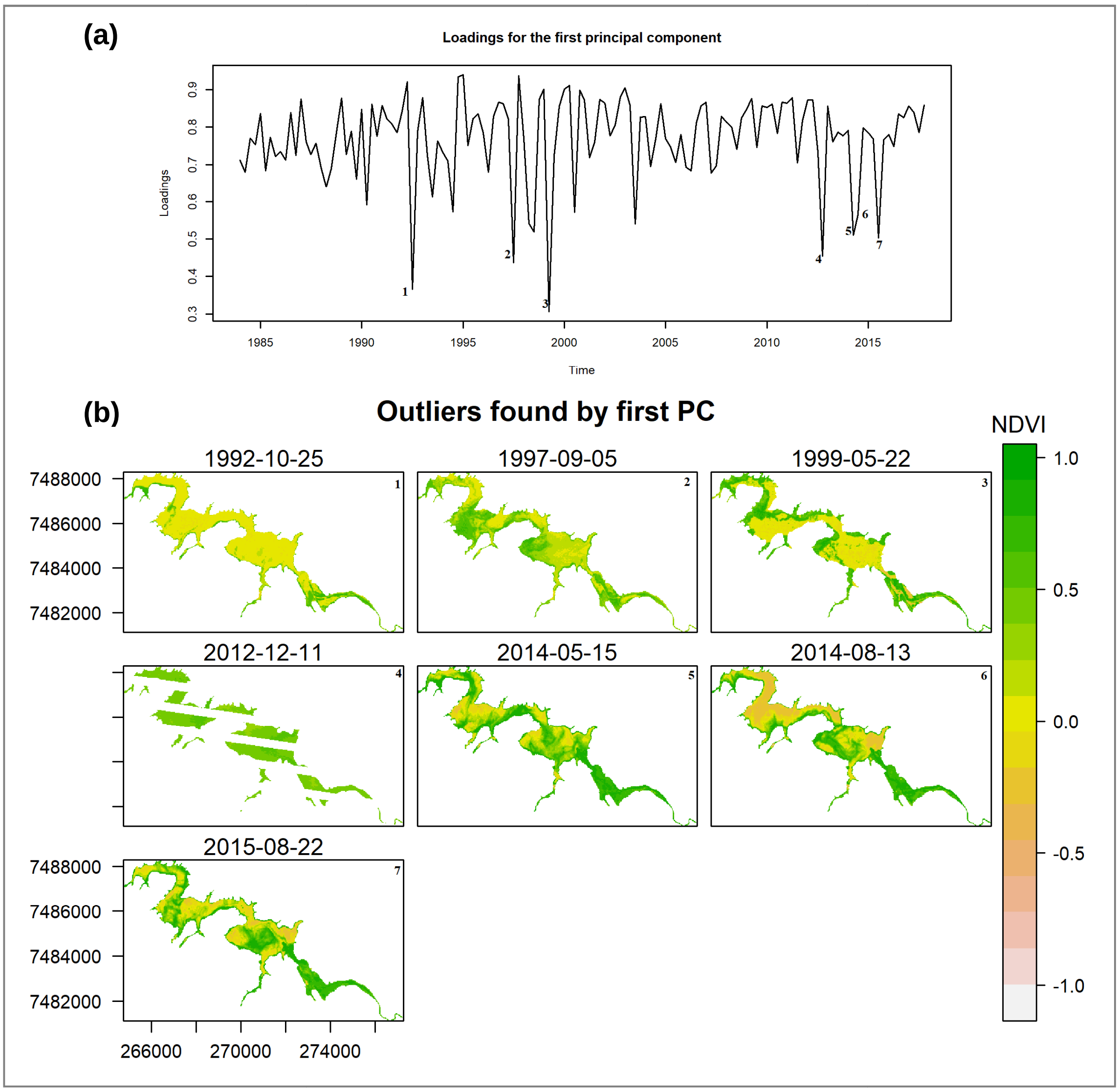


Figure 16. Temporal profile of the PC1 loadings highlighting (a) the outliers, and (b), the images corresponding to these outliers. PC1 represents 61% of variability of NDVI data.

While the spatial distribution of the NDVI temporal means and variances indicate the recurrent presence of macrophytes in sectors designated as A and B, the NDVI images highlighted as outliers in the PC1 loadings profile show scenarios in which the spatial distributions of the macrophytes in the reservoir were quite different. The first NDVI image (10-25-1992) showed a small incidence of macrophytes in the reservoir, different from the patterns observed on most dates, which exhibited a high density of plants dispersed differently on the reservoir. The low temporal correlation that differentiated the NDVI images obtained on 1997-09-05, 1999-05-22, 2014-05-15, 2014-08-13 and 2015-08-22 is probably associated with the environmental or meteorological variables that influenced the spread of macrophytes. The exception observed was in the image of 2012-12-11, a scene acquired by the ETM+

sensor that shows strips without data due to SLC-Off. Then, in the case of missing values in the SLC-Off pixels, they were interpolated using a Kalman filter applied to each time series once they were extracted for each pixel.

4.2.3 Regions of homogeneous temporal variability

The definition of the specific sites to monitor the trend behavior, seasonality and ruptures in the temporal pattern in the spatial dispersion of macrophytes in the reservoir, considered the mapping of the regions that presented similar temporal correlations. Figure 19(a) shows the spatial distribution of PC1 scores, expressed in number of standard deviations around the mean, while Figure 19(b) shows the regions formed by a K-means clustering of the similar PC1 scores.

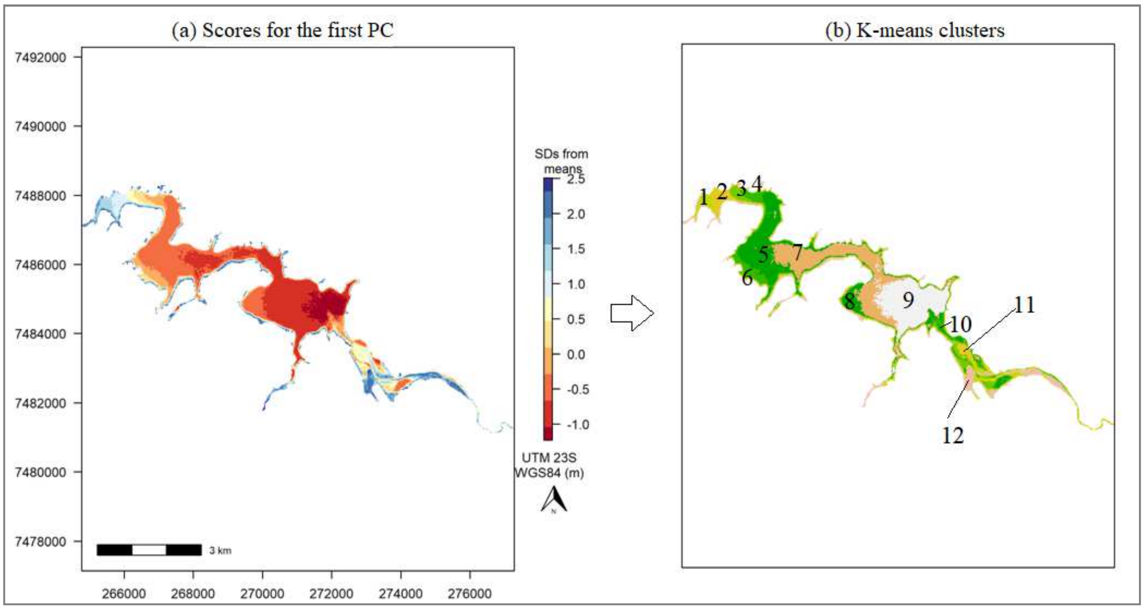


Figure 17. Spatial distribution of temporal PC1 scores (a) expressed in number of standard deviation, and (b), regions of similar PC1 scores defined by K-means clustering.

Figure 19(a) represents the spatial distribution based on the results of the PC1, showing similar patterns observed in results of Figure 17(a). Since most of the correlations of PC loadings of Figure 18(a) are equal or higher of 0.8, it can be concluded that more than 64% of variance of the area are explained. These correlations show that their respective quarters have a strong similarity to the one depicted in Figure 19(a).

The PC1 scores clustering by the K-means algorithm with eight classes shows the temporal variability in the occurrence of macrophytes for each pixel of the image.

This clustering was compared with Figure 17(a) where it was verified that the sectors A and B were segmented in small regions, just as the vast areas in the middle of the reservoir defined large regions characterized by small variations over time. In addition, the segmented regions can be spatially significant, in terms of the macrophytes occurrence in the reservoir, so those groups of pixels included in the same class without spatial connection among them define different regions. Thus, the twelve sectors depicted in Figure 19(b) were considered representative of the temporal variability in the spatial dispersion of macrophytes in the Salto Grande reservoir.

4.2.4 Trend and temporal changes in homogeneous areas

As specified in the methodology, the analysis of trend and seasonality, as well as the breaks in the time series, considered regions of homogeneous temporal variability, spatially aggregated by the K-means algorithm. To characterize these regions in the reservoir context, they were grouped according to their position in the water body. The results of the NDVI time series decomposition by the BFAST algorithm for each region are presented according to this grouping. Thus, sectors 1, 2 and 3 that define the location near the dam are shown in Figure 20, while Figure 21 depicts segments 4, 5, 6 located upstream of the dam. Sectors 7, 8 and 9, in the middle of the reservoir, are represented in Figure 22 while sectors 10, 11 and 12 that are under the river influence and located upstream of the reservoir are illustrated in Figure 23.

Thus, Figures 20, 21, 22 and 23 present plots, each one referring to a specific sector, plus a map indicating the location of the sector in the reservoir context. The graphs comprise three horizontal panels, the upper one shows the average values of monthly NDVIs over time (Y_i) and the other panels indicate the results of the time series decomposition by the BFAST for each region considered. S_t is the seasonality component, and T_t shows the breakpoints detected in the general trend (where ρ is the slope and p is p-value). The remainder component is omitted.

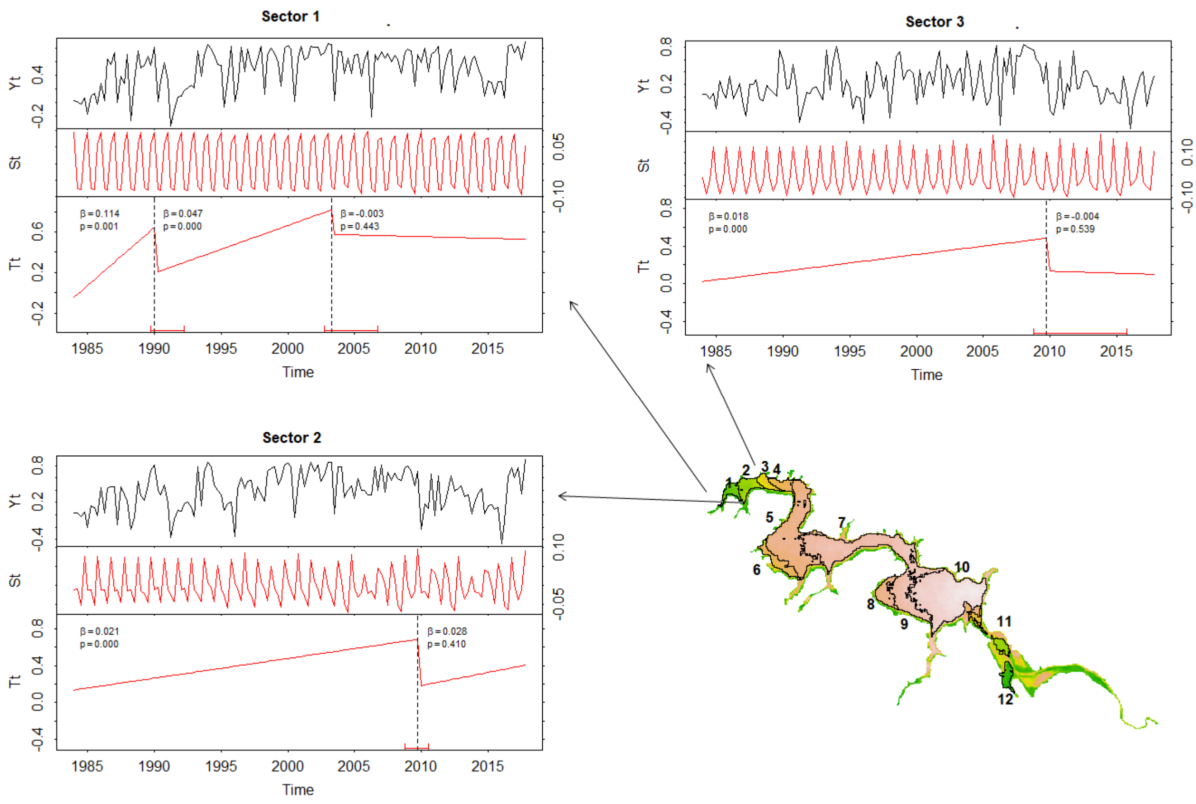


Figure 18. Time series generated for sectors 1, 2 and 3 (near dam) and their decomposition by BFAST. For each graph, Y_t represents the average values of monthly NDVIs over time, S_t is the seasonality component, T_t is the breakpoints detected in the general trend (where β is the slope and p is p-value).

In the Figure 20, region 1 shows two significant breakpoints, one in decade of 1990 and the other near 2005, with negative magnitudes. According to CETESB (1995), while there exists no studied sample point for that epoch in Salto Grande, Atibaia River contained above average quantities of ammoniac nitrogen and other nutrients, which also contribute for eutrophication. In the decades of 80s and 90s, also, the control was made by using “drag line” tractors to remove the aquatic plants and deposit them directly on dump trucks (CPFL, 2009). The 2004 – 2010 results of Macedo (2011) indicated this period as having constant macrophytes near the dam, what is shown in the constant trend component of Figure 20 (sector 1, i.e., insignificant at 0.05 level). Moreover, according to Macedo (2011), 2010 was the period with less quantity of aquatic plants in the areas of the reservoir, thus corroborating our results in charts 1 and 3 of Figure 20, which suggest decreasing magnitudes near that year.

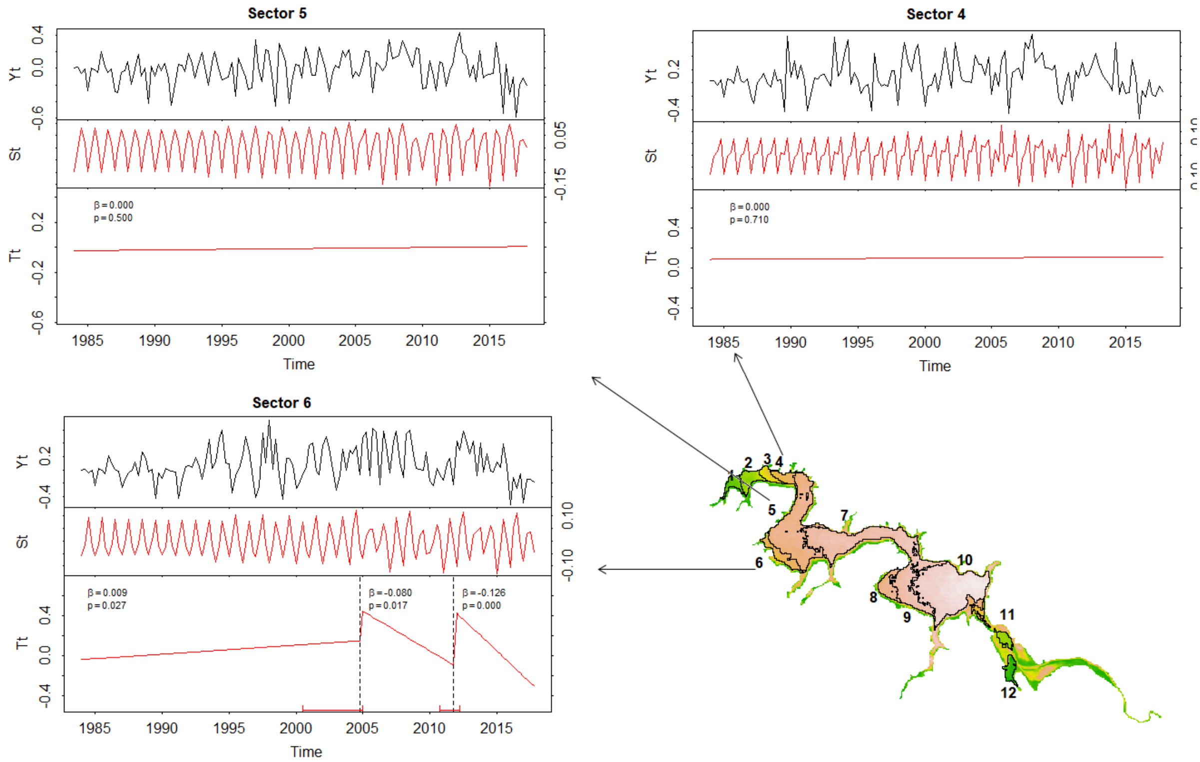


Figure 19. Time series generated for sectors 4, 5 and 6 (upstream of the dam) and their decomposition by BFAST. For each graph, Y_t represents the average values of monthly NDVIs over time, S_t is the seasonality component, T_t is the breakpoints detected in the general trend (where β is the slope and p is p-value).

From Figure 21, charts 4 and 5 present no breakpoints and trends are almost constant over time, while chart 6 presents two breakpoints. The first two clusters represent low NDVI values, mostly between -0.2 and 0.2, indicating the absence or browning of the macrophytes, probably due to wind flux (Martins et al., 2011). It also confirms the spatial pattern of low correlation observed in the PC1 scores (Figure 19a) for this region. Furthermore, chart 6 indicated two breakpoints followed by a decrease in the trend component in agreement with Macedo (2011), whose results showed 2010 as the year with fewer amounts of macrophytes.

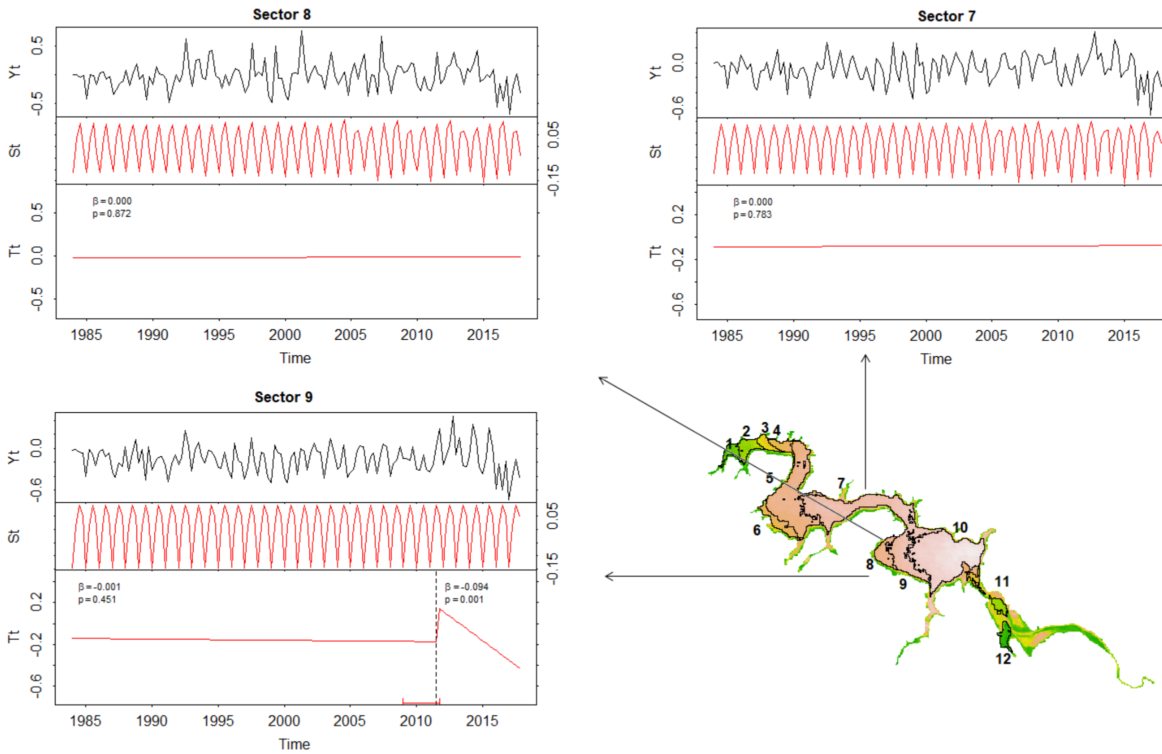


Figure 20. Time series generated for sectors 7, 8 and 9 (middle of the reservoir) and their decomposition by BFAST. For each graph, Y_t represents the average values of monthly NDVIs over time, S_t is the seasonality component, T_t is the breakpoints detected in the general trend (where β is the slope and p is p-value).

From charts 7, 8 and 9 of Figure 22, no abrupt changes for the first two sectors in the trend components neither significant slopes (at 0.05 level) were observed. These regions also represent low NDVI values (around 0), indicating the absence of aquatic plants, even after the breakpoint in Sector 9. These areas are located in the middle of the reservoir, where the initial analysis showed low variability in addition to lower mean values for the period analyzed (Figure 17).

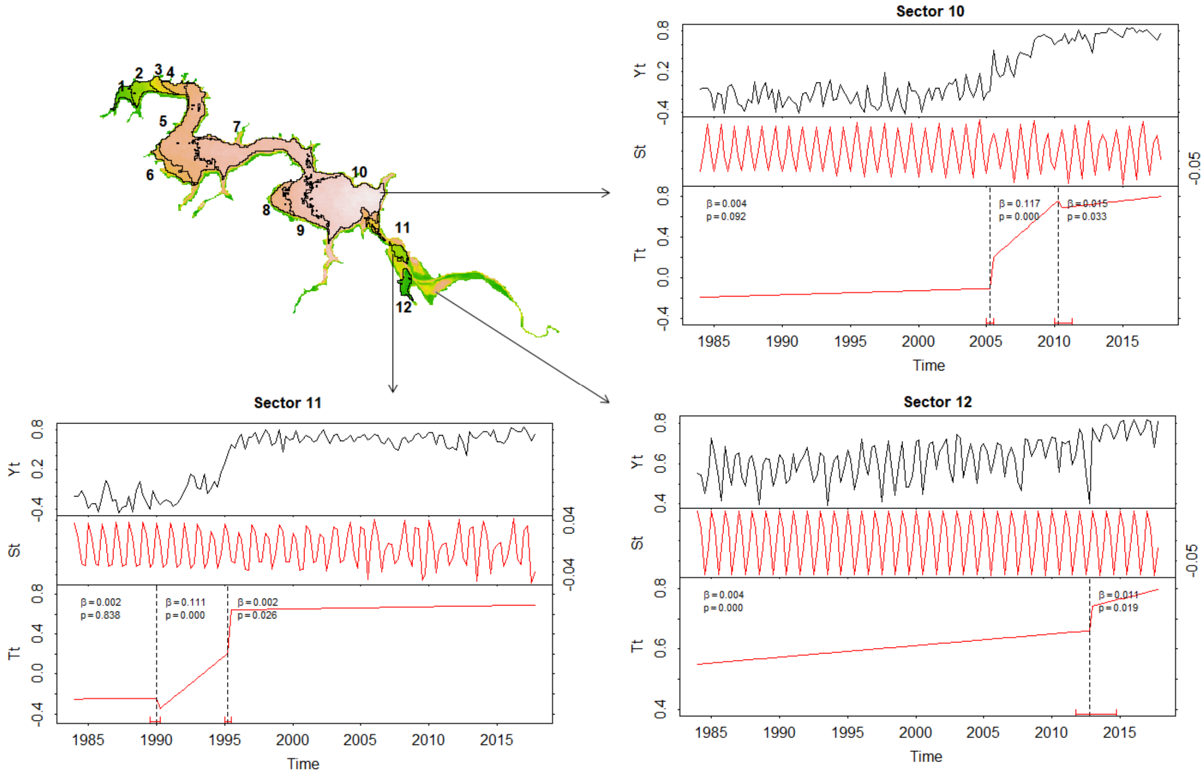


Figure 21. Time series generated for sectors 10, 11 and 12 (upstream of the reservoir) and their decomposition by BFAST. For each graph, Y_t represents the average values of monthly NDVIs over time, S_t is the seasonality component, T_t is the breakpoints detected in the general trend (where β is the slope and p is p-value).

For area B of the reservoir, where sectors 10, 11 and 12 are located (Figure 23), most variability and high NDVI values were observed. On the one hand, for these three regions, for the entire study period, there is almost no decrease in NDVI, except for years 1990 (chart 11) and 2010 (charts 10 and 12), where a smooth decaying was observed in the trend component. On the other hand, values of NDVI were higher than 0.4 since 2010, indicating high presence of plants. These charts presented more prominent positive magnitudes indicating a rapid increase of NDVI values than negative ones. According to CETESB (2005), in 2010, high inputs of phosphorus were found near Atibaia River which contributed to the proliferation of bacteria and plants that were recurrent during the following years. This is probably due to agricultural activities near the region (especially in summer) and the effluent from domestic sewers originating from Campinas and nearby municipalities (Fonseca (2013) and Carvalho et al. (2018)).

As can be seen from BFAST results, areas A and B had several abrupt (negative and positive) changes (Figures 20 and 23), while most part of the middle of reservoir did not present any abrupt change (Figure 22), in agreement with Figure 17. At least one breakpoint was found in time series for each pixel where magnitude differs from zero, and also indicating that at least one abrupt change occurred in these areas as well as the gradual changes. Area B had the biggest cluster of abrupt changes (both negative and positive), indicating that almost the entire area suffered changes, i.e., only few pixels presented a value of zero magnitude, explaining also the great variability of NDVI found in the region.

4.3 Analysis of the influence of exogenous variables on the spatiotemporal behavior of aquatic macrophytes in the Salto Grande Reservoir

By using the PCA and K-means classification results of Experiment 2 as seen in Figure 19(b), the mean NDVI of the pixels within each sector was calculated and the mean NDVI time series were built again. NDVI for the entire reservoir area was also calculated (referred as global NDVI). Finally, groups of interest (regions) were created as follows: 1, 2 and 3 were considered the "near dam" region. Sectors 4, 5, 6, 7, 8 and 9 as "reservoir body" region and 10, 11, 12 sectors were the "near the river" region. Figure 24 summarizes the sectors and areas defined in this study.

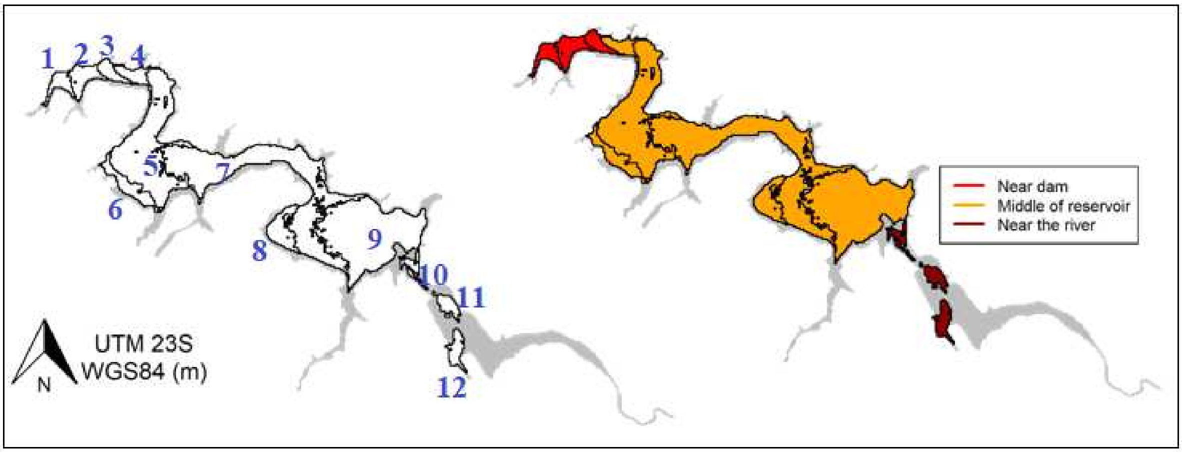


Figure 22. Twelve sectors obtained by K-means clustering of PC1 scores, aggregated into three regions: near dam, middle of reservoir and near the river.

The features extracted from the images that characterized the reservoir and its surroundings were: the NDVI values for each pixel of the 12 segmented areas (grouped into three sectors in the analysis of the influence of climatic variables - Figure 24) and the average NDBI and NDVI calculated for the reservoir environment (Figure 14(b)) and the water body (Figure 14(c)), respectively, in the analysis of the influence of the environment.

4.3.1 Anthropogenic influence in the spatial-temporal macrophytes dynamics

The BFAST decomposition of the mean NDVI time series for the entire reservoir, obtained by using mask from Figure 14(c) for each quarterly image and the mean NDBI time series for the surroundings of the reservoir (mask from Figure 14(b)), are shown in Figure 25. The graph comprises four horizontal panels: the upper one shows the mean values of NDVI (in red) or NDBI (in blue) over time (Y_t) and the other panels indicate the results of the time series decomposition by the BFAST for each index. S_t is the seasonality component, T_t is the general trend (with a single breakpoint in the NDBI trend) and e_t is the remainder component.

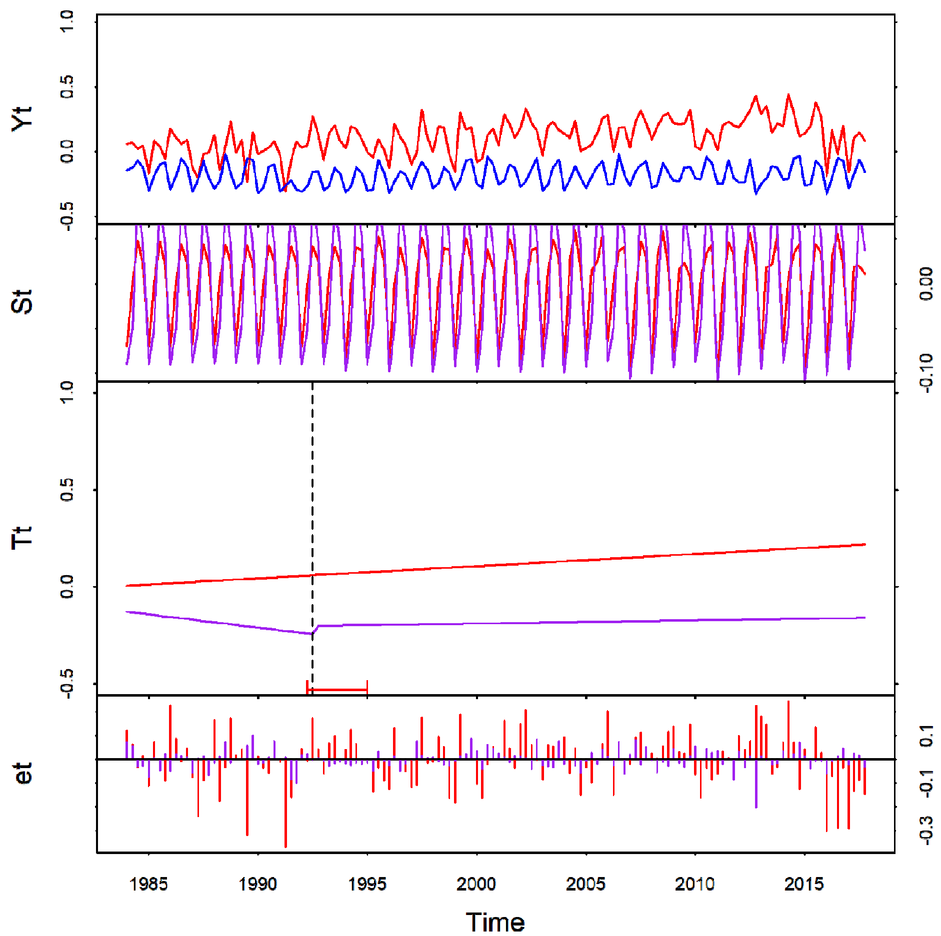


Figure 23. NDVI time series (in red) for reservoir and NDBI time series (in blue) for its surroundings, along with BFAST decomposition. Y_t represents the mean values of NDVI or NDBI over time; S_t is the seasonality component, T_t is the general trend.

By extracting trend and seasonality of global NDVI and each climatic variable (Air Temperature, Meridional Wind, Zonal Wind and Precipitable Water) using the BFAST algorithm, no breakpoint was found in the global NDVI time series. However, an increasing trend was observed in the studied period: mean NDBI time series only presents an increasing trend after the breakpoint in 1992. Comparing the time period where both NDVI and NDBI increases in trend, it is observed that all values of NDVI are positive, while the NDBI values are negative. Using the convention, trends in NDBI showed continuous predominance of vegetation more than urban areas or bare soil in the region.

The mean values represented in the original NDVI and NDBI time profiles make up the first panel (Y_t) of Figure 25 and pointed to the higher variability of NDVI, associated with the occurrence of macrophytes in the reservoir. The mean NDBI time series resulted in low values, probably due to the contribution of bare soil in the agricultural area. Seasonal fluctuations in NDVI and NDBI followed the same pattern, as shown by the seasonality component (S_t). This suggests that the periods of higher occurrence of macrophytes in the reservoir are synchronized with the increase of bare soil in its surroundings, as the larger variations in NDVI values (the residuals in the bottom panel (e_t)) confirm.

The T_t component showed a growing tendency in the macrophytes occurrence in the reservoir (increase in NDVI), while NDBI remained practically constant after a period of decline between 1984 and 1992. In this initial period, the effects of the water quality degradation were already noticed, triggered by the intense urbanization that had its peak in the early 1980s (Martins et al. 2011). The installation of industrial parks and the spread of sugarcane cultivation seem to have led to stabilization of the anthropogenic occupation. However, in the water body as a whole, this process triggered the development of extensive aquatic macrophyte banks, accumulating on the margins and near the dam displaced by the flow of water and wind.

The cross-correlation between the mean NDVI time series of masks 14(c) and the mean NDBI time series of mask 14(b)-applied scene was calculated. The value obtained was 0.316, meaning that they have slightly positive correlation and significant at 5% level.

4.3.2 Climatic variables influence on spatiotemporal dynamics of macrophytes

From the data provided by NOAA/ESRL, the time series were constructed for air temperature, southern meridional winds, zonal winds and precipitable water. Figure 26 shows these climatic variables time series (Y_t) as well as the seasonality (S_t) and trend (T_t) components resulting from the application of BFAST.

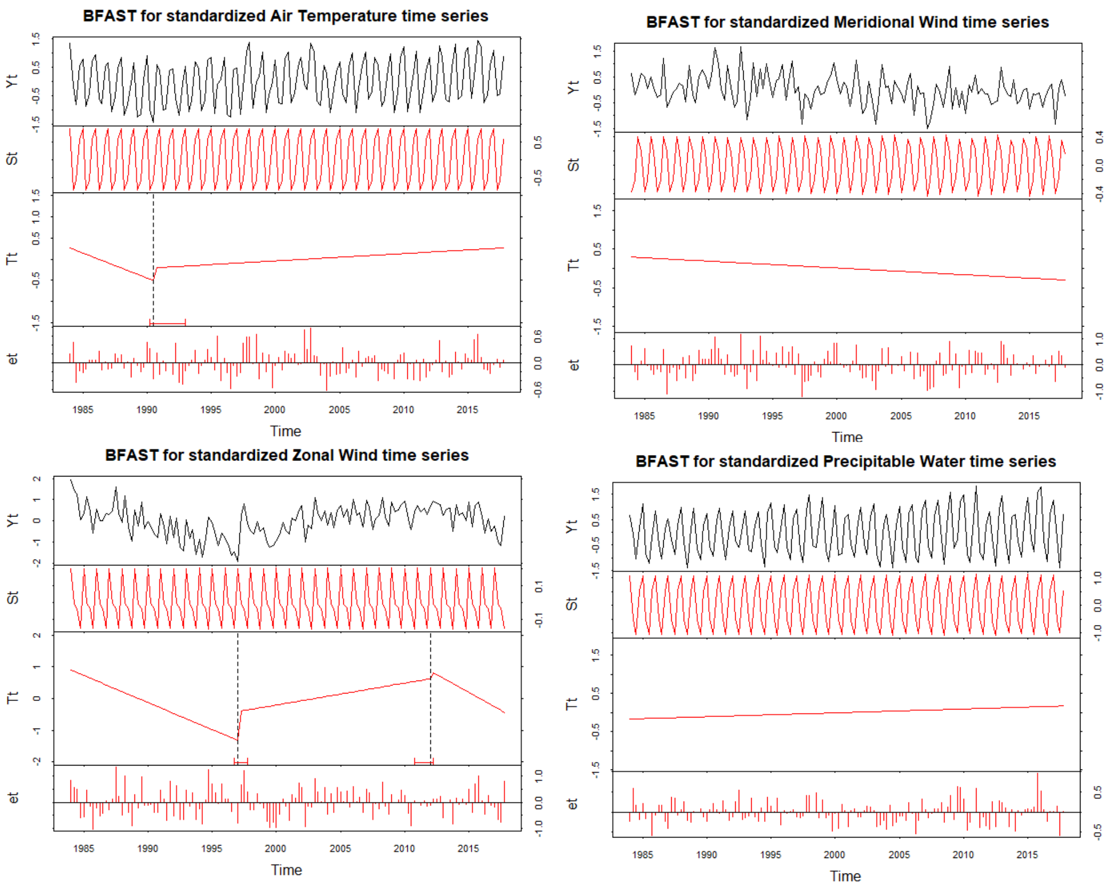


Figure 24. Standardized climatic time series and their seasonality (S_t) and trend (T_t) components resulting from the application of BFAST procedure to: air temperature, meridional winds, zonal winds, and precipitable water.

The marked cyclical characteristics of the temporal variations of Air Temperature and Precipitable Water is noticeable even in the original profiles of these variables, but the seasonality component captured this cycle in the meridional and zonal winds data. Regarding the general behavior of these variables over time, meridional winds presented a decreasing trend without breakpoints, while precipitable water presented an increasing trend also with no breakpoints. Air Temperature showed a trend breakpoint in 1990 and zonal winds, however, presented two breakpoints in its trend component: in 1997 and in 2012. The periods of 1990-1993 and 1997-1998 were periods of strong El Niño occurrence and 2012 was period of La Niña (Sá et al., 2018), thus, breakpoints found on air temperature may indicate the behaviour on El Niño, with increase in air temperature. El Niño and La Ninã may also have influence on zonal winds breakpoints, which are more forceful eastward-blowing winds in the tropical western areas and weakened easterly winds near the Americas (Carlowicz and Schollaert, 2017).

4.3.2.1 NDVI and climatic variables correlation

The cross-correlation of Y_t, i.e., between the mean NDVI time series of mask from Figure 14(c)-applied scenes and the climatic variables time series were also obtained at 5% significance level. Correlating the global NDVI with Air Temperature, Zonal and Meridional Winds, and precipitation resulted in the correlation values of -0.07, 0.07, 0.04 and -0.2, respectively, indicating that only precipitation is slightly negatively correlated with the global NDVI.

On the other hand, the correlations between the respective trend and seasonality time series components of each climatic variable with the global NDVI result in high correlations for some of the variables, as shown in Table 12, where trend component is found in the lower part of principal diagonal and seasonality for upper principal diagonal.

Table 12. Correlation matrix obtained for trend components (values in blue above the diagonal) and seasonality (values in green above the diagonal) extracted from the time series of the global NDVI time series and each of the climatic variables: air temperature (Air Temp.), southern and zonal winds (Mer. Winds and Zon. Winds), and precipitable water (Prec. Water).

	NDVI	*Air Temp.*	*Mer. Wind*	*Zon. Wind*	*Prec. Water*
NDVI	1	**-0.5958**	**0.8751**	**-0.8835**	**-0.7551**
Air Temp.	**0.7061**	1	**-0.3217**	**0.3092**	**0.93**
Mer. Wind	**-1**	**-0.7061**	1	**-0.7659**	**-0.6191**
Zon. Wind	**0.1947**	**0.5084**	**-0.1947**	1	**0.4328**
Prec. Water	**1**	**0.7061**	**-1**	**0.1947**	1

By considering global NDVI trends, they are strongly correlated with the climatic trends of the reservoir area, with values of -1 and 1 also being found, i.e., the piecewise regression lines for those trend components were similar. For seasonality, high positive and negative values were also observed (0.8751 and -0.8835) when comparing NDVI seasonality with the climatic variables seasonality.

4.3.2.2 Influence of climatic variables on macrophytes occurrence in specific reservoir sectors

To analyze the influence of climatic variables on the occurrence of macrophytes in the different regions of the reservoir over time, the correlation between the time series of these variables and the mean NDVI from each of the twelve clusters resulting from the K-means applied to the first principal component was calculated. The time series cross-correlation was calculated and presented in Figure 27 by means of a bar plot, with each bar representing sequentially sectors 1 to 12, grouped in regions "near dam", "reservoir body" and "near river" by colors.

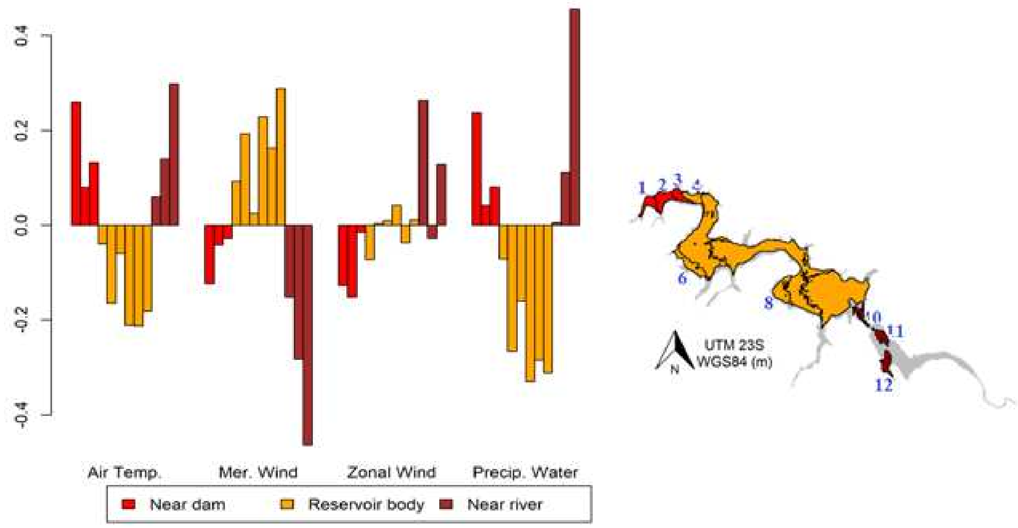


Figure 25. Correlation between climatic variables (Air Tempetature, Meridional and Zonal Winds and Precipitable Water) and the mean NDVI of each area of reservoirs defined by a temporal PCA. The bars represent sequentially sectors 1 to 12, grouped in regions by colors.

The correlation values between the time series of the climatic variables and the NDVI by area were all lesser than 0.5, but all, except zonal winds, defined a different behavior for the three regions of the reservoir. For air temperature and precipitable water, similar behaviors of correlations were observed in Figure 27: they were positively correlated with NDVI in the near dam and near river sectors and negatively correlated with NDVI in the reservoir body. Meridional winds were negatively correlated with NDVI in these sectors, while positively correlated in the reservoir body. Zonal winds seem to be uncorrelated with the reservoir body, but negatively in the near dam and mostly positively in the near river. The highest negative and positive correlations occured for meridional wind and precipitation, respectively, both in the near river sector. Aside from zonal wind variable, all others produced clear correlations patterns.

By decomposing each time series using BFAST into trend and seasonality and correlating the results with standardized mean NDVI, the followings values on Figures 28 and 29 were obtained for the three sectors and areas, with lines representing sectors 1 to 12, sequentially.

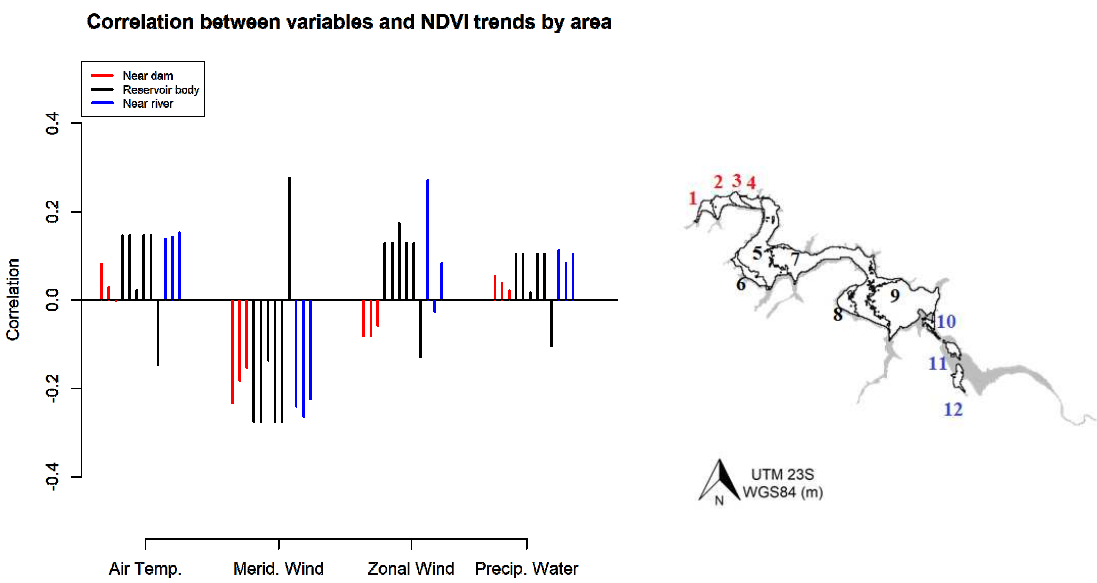


Figure 26. Correlation between the trend components of the time series of each climatic variable and the mean NDVI for the segmented sectors grouped by region of the reservoir. The vertical lines represent the correlations for sectors 1 to 12, sequentially.

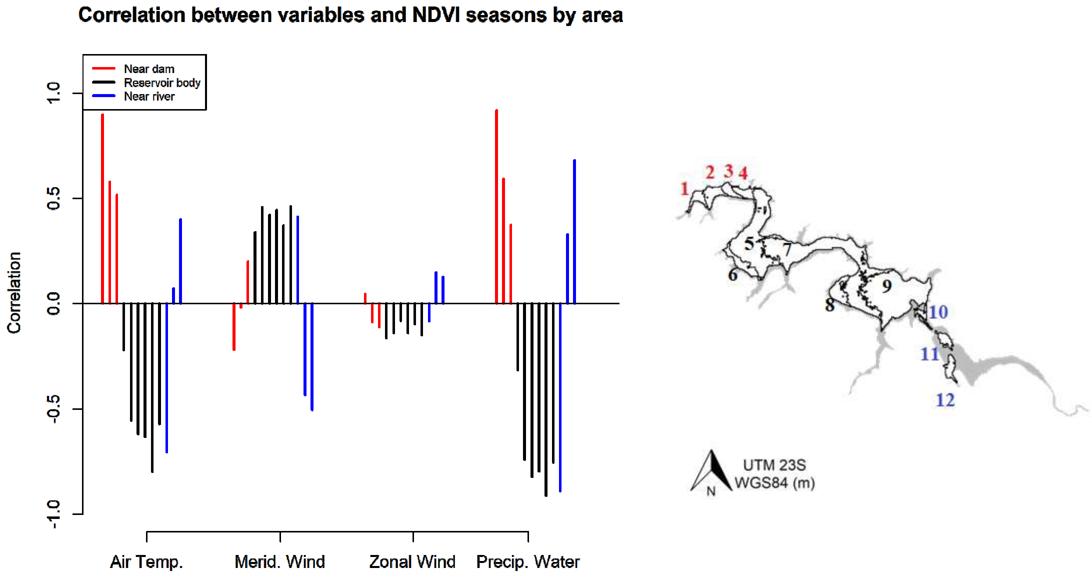


Figure 27. Correlation between the seasonality components of the time series of each climatic variable and the mean NDVI for the segmented sectors grouped by regions of the reservoir. The vertical lines represent the correlations for sectors 1 to 12, sequentially.

The trend component over time resulted in a low correlation between the mean NDVI of the segmented areas and the climatic variables, none of which exceeded 0.3, as shown in Figure 27. On the other hand, the greater correlations obtained between the seasonality components of the time series of the climatic variables and the mean NDVI of the segmented areas (Figure 29) indicate that the spatial dispersion and persistence of aquatic macrophytes in specific locations of the reservoir are related to the cyclical variations of precipitation, air temperature and winds from the South. Both precipitation and temperature, variables that defined a high correlation with the NDVI for the seasonality component, had similar behavior for the three sectors of the reservoir: positive correlations in the sectors near the dam and in the environment under the influence of the river. However, the seasonal variability in the occurrence of macrophytes in the reservoir body had a negative correlation with these variables (all larger than 0.5, with the exception of area 4).

Meridional winds had a positive correlation and approximately 0.5 with the areas defined as the reservoir body, indicating that the occurrence of macrophyte banks in this sector is influenced by the seasonal behavior of these winds. Zonal winds presented very little influence in all sectors.

4.3.3 Granger-causality models

By the results of previous sections, a study of the causality of the variables air temperature, zonal winds and precipitable water in the global NDVI was carried out. No single root presence was detected; thus, no differentiation was necessary for these time series. Granger's causality was applied to the time series using the fourth lag of each climatic variable. The models were defined by adding a lag in each variable until P(<F) <0.05, that is, considering a significance level of 5%. The climatic variables considered the number of lags used and the p-value obtained for the models that met the 5% significance level, are shown in Table 13. The variables that defined each model are shown in equations (11) to (14).

Table 13. Granger’s causalities (GC) p-values between the environmental variables and global NDVI for maximum lag obtained by BIC.

Cause	*Lags used*	*Variable GC NDVI P(<F)*
Air temperature	1, 2, 3, 4	0.0001
Precipitable water	1, 2, 3, 4	0.0086
Meridional wind	1	0.0006
All three	1, 2	0.0001

Model equations (11) through (14) are the full (unrestricted) models, which were compared with their respective restricted model, i.e., the models obtained by setting all $\beta s, \gamma s$ and δs all equal 0 for each equation. The column Variable GC NDVI shows all p-values below the significance value 0.05, indicating that all climatic variables Granger-cause NDVI in all models from (11) to (14). This means that:

(i) the air temperature time series Granger-causes NDVI, as the knowledge of past values of air temperature (in this case, four quarters behind) is useful (as pointed by the p-value < 0.05) in explaining the future of NDVI state over and above knowledge of the past history of NDVI;

(ii) the precipitable water time series Granger-causes NDVI, since knowledge of the past four values of precipitable water helps in explaining the future of NDVI state as same in (i);

(iii) the meridional wind time series granger-causes NDVI, this time with only the knowledge of only one lag (i. e., last quarter), helping in predicting future NDVI as well;

(iv) with a model using the three variables air temperature, precipitable water and meridional winds, by using only two lags of each variable, it is possible to say that they Cranger-cause NDVI. This is useful as correlation does not imply causation.

5 CONCLUSIONS AND RECOMMENDATIONS

Monitoring the growth of macrophytes in aquatic environments is essential to establish management practices and control their over-abundance, preserving the sustainable conditions of the ecosystem and the health of the population that lives close to them. As demonstrated in this study, remote sensing, together with historical data of climatic variables, statistically analyzed through time series, can provide information on growth trends and spatiotemporal behavior of aquatic macrophytes in eutrophic lakes and reservoirs.

The Experiment 1 study with Pixel QA band enabled to compare the band values for entire images and for the same images clipped for a small region. In the case study, 0.17% of the entire image was analysed and was possible to note the differences: the clipped images showed reduction of QA values and most of them showed reduced standard deviation of remained component. Considering the full image time series, using the mean values as a QA indicator should be assessed, for example, with the information of F-mask algorithm, since 40% of the entire images were considered as low confidence cloud shadow presence. This indicates that a possible metric could be created based on these values and values of standard deviations for small areas, for example, showing a user which images are considered without clouds (cloud shadows) with F-mask algorithm, which got a small mean QA value, small values of standard deviation, change of confidence level, etc. Furthermore, this first experiment showed that the images have the necessary quality to precede with the time series analyses.

The exploratory analysis of mean and variance in Experiment 2 already highlighted the persistence of macrophytes behavior in the reservoir. Applying PCA and adopting PC1 results for K-means clustering made it possible to map sectors of homogeneous NDVI variability. By using the BFAST algorithm, it was possible to analyze the aggregated NDVI time series for each sector and to identify macrophytes persistence. Most abrupt changes, found by the breakpoints, also highlighted the importance of monitoring regions of the dam and near the river. From all BFAST charts of Experiment 2, it is important to note that most of the slopes were positive and significant at 0.05 confidence level, indicating increasing of NDVI through years. Even after a breakpoint showing decrease (like in sector 1 in Figure 18), the trend started to grow positively again. Interestingly, only chart 6 had significant decrease in NDVI in recent years (after two abrupt changes of increase). Results presented in this paper showed that BFAST applied to the K-means clustered areas of reservoir was capable of identifying the reservoir heterogeneity and abrupt breaks in similar periods in more than one region and an increasing of macrophytes around the almost entire reservoir.

In Experiment 3, to evaluate the influence of exogenous variables on the dispersion temporal changes in the macrophyte occurrence in reservoirs located in anthropic areas, a time series analysis approach based on attributes extracted from Landsat images and climatic data was designed to investigate two scenarios. Firstly, the influence of anthropogenic occupation around the reservoir was considered, comparing NDVI and the NDBI time series. The influence of weather conditions was evaluated through joint analysis of NDVI time series and climatic oscillation time profiles, constructed from the NOAA historical records of the air temperature, meridional and zonal winds and precipitable water. The analysis developed with the information provided by the indices and exogenous variables within a time-series decomposition approach highlighted important patterns and the utility of these variables in the macrophytes spatiotemporal behavior. With a higher variability, NDVI showed similar seasonal patterns with NDBI while maintaining an increase in trend for all studied period, confirming that the macrophytes in the reservoir still are in greening process. Moreover, this process is also highlighted when not only trend and seasonality are correlated in specific locations of the area with the climatic variables (precipitable water, air temperature and meridional winds clearly showed patterns with NDVI), but also when incorporating these three variables to the NDVI series through a Granger-causality model. When combined to lagged values, the climatic variables were found to Granger-causality NDVI, thus, important to describe and predict the macrophytes' persistence.

The developed method in this research, based on the analysis of time series generated from free of charge Landsat images and climatic data, can be useful in the management of reservoirs subject to systematic macrophyte infestations, both for government agencies and private companies. Knowing when and where control measures should be applied makes the process more efficient and allows choosing the most appropriate type of management for each situation. Limitations were also found, mainly the lack of cloud-free images and correct images of Landsat 7 ETM+ Slc-Off period. The quarterly time series approach was necessary in order to get equispaced time series and interpolation for missing values and correction of Slc-Off problem.

As recommendations, the entire methodology (time series construction using the first principal component to analyze the quality of image, segmentation in homogeneous zones, statistical tools like BFAST decomposition, Granger-causality, and inspection of QA band) can be applied and adjusted for crop monitoring, as well as productivity breaks, forested areas, etc. Going further, the quality assessment is also possible, being possible to investigate new metrics of quality which can be created by comparing, for example, results of Experiment 3 and scores generated for each user's action through their researches. Furthermore, as pointed in Experiment 2, using the first principal component loadings could also help in the analysis, as images with low quality were highlighted during the process.

REFERENCES

ADAMI, M., BERNARDES, S., ARAI, E., FREITAS, R. M., SHIMABUKURO, Y. E., ESPÍRITO-SANTO, F. D. B., RUDROFF, B. F. T., ANDERSON L. O., Seasonality of vegetation types of South America depicted by moderate resolution imaging spectroradiometer (MODIS) time series. International Journal of Applied Earth Observation and Geoinformation, v. 69, 2018. DOI: 10.1016/j.jag.2018.02.010

ALQURASHI, A. F., KUMAR, L. Investigating the use of remote sensing and GIS techniques to detect land use and land cover change: A review. Advances in Remote Sensing**,** v. 2, n. 02, p. 193, 2013. DOI: 10.4236/ars.2013.22022

ARVOR, D., DAHER, F. R. G.., BRIAND, D., DUFOUR, S., ROLLET, A-J., SIMÕES, M., FERRAZ, R. P.D., Monitoring thirty years of small water reservoirs proliferation in the southern Amazon with Landsat time series. ISPRS J. Photogram. Remote Sensing, 2018. doi:10.1016/j.isprsjprs.2018.03.015

ATTANASIO, A., PASINI, A., TRIACCA, U., Granger causality analyses for climatic attribution. Atmospheric and Climate Sciences, v. 3, p. 515 – 522, 2013.

AVISSE, N., TILMANT, A., MÜLLER, M. F., ZHANG, H., Monitoring small reservoirs' storage with satellite remote sensing in inaccessible areas. Hydrol. Earth Syst. Sci, v. 21, 2017 DOI:10.5194/hess-21-6445-2017.

AWANGE, J. L. ONG´ANG´A, O., Lake Victoria: Ecology, environment, resources, Springer, 2005.

BADLANI, B., PATEL, A. N., PATEL, K., KALUBARME, M. H., Urban growth monitoring using remote sensing and Geo-informatics: case study of Gandhinagar, Gujarat State (India). International Journal of Geosciences, v. 8, p. 563 – 576, 2017. DOI: 10.4236/ijg.2017.84030

BAI, J. PERRON, P. Computation and analysis of multiple structural change models. J. Appl. Econometrics, 2003. DOI: 10.1002/jae.659

BERGIER, I., ASSINE, M. L., McGLUE, M. M., ALHO, C. J. R., SILVA, A., GUERREIRO, R. L., CARVALHO, J. C., (2018). Amazon rainforest modulation of water security in Pantanal wetland. Science of Total Environment, v. 619, p. 1116 – 1125. DOI:j.scitotenv.2017.11.163

CARLOWICZ, M., SCHOLLAERT, S., El Niño, 2017. Available in <http://earthobservatory.nasa.gov/features/ElNino>.

CARVALHO, J. C. B., MATIAS, L. F., CAPORUSSO, D., Monitoramento da qualidade ambiental no reservatório de Salto Grande, Americana-SP, 2018.

CETESB. 1995. Relatório de qualidade das águas interiors do estado de São Paulo. Séries Relatórios, São Paulo, CETESB, 1995. Available in <http://www.cetesb.sp.gov.br>.

CETESB. 2007. Relatório de qualidade das águas interiors do estado de São Paulo. Séries Relatórios, São Paulo, CETESB, 2010. Available in <http://www.cetesb.sp.gov.br>.

CHAVEZ JR., P. S., An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sensing of Environment, v. 24, p. 459 – 479, 1988. DOI: 10.1016/0034-4257(88)90019-3

CHAVEZ JR., P. S., Image-based atmospheric corrections – revisited and improved. U. S. Geological Survey, v. 62, no. 9, p. 1025 – 1036, 1996.

CHU, C., HORNIK, K., KUAN, C., MOSUM tests for parameter constancy. Biometrila, v. 82, no. 3, p. 603 – 617, 1995. DOI: 10.2307/2337537

COMMANDEUR, J. J. F., KOOPMAN, J., An introduction to State space time series analysis. Practical econometrics, New York, Oxford, 2007.

CONFORTE, J. C., ROBALLO, S. T., FERREIRA, N. J. Variabilidade temporal do fluxo da radiação de ondas longas no Brasil e vizinhanças no período de 1979 a 2008 – Parte II: Ciclo anual. Anais XV Simpósio Brasileiro de Sensoriamento Remoto – SBSR, Curitiba, PR, p. 2203, 2011.

CPFL Energia. 2009. Controle de plantas aquáticas no reservatório de Americana.

DAR, N. A., PANDIT, A. K., & GANAI, B. A. Factors affecting the distribution patterns of aquatic macrophytes. *Limnol. Rev.*, v. 14, n. 2, p. 75-81, 2014. DOI: 10.2478/limre-2014-0008

DARMAWAN, Y., SOFAN, P., Comparison of the vegetation indices to detect the tropical rain forest changes using breaks for additive seasonal and trend (BFAST) model. International Journal of Remote Sensing and Earth Science, v. 9, no 1, p 21 – 34, 2012. DOI: 10.30536/j.ijreses.2012.v9.a1823

DE JONG, R., BRUIN, S., WIT, A., SCHAEPMAN, M. E., DENT, D. L., Analysis of monotonic greening and browning trends from global NDVI time-series. Remote Sensing of Environment, v. 115, n. 2, p. 692-702, 2011. DOI: 10.1016/j.rse.2010.10.011

DE JONG, R., VERBESSELT, J., ZEILEIS, A., SCHAEPMAN, M., Shifts in global vegetation activity trends. Remote Sensing, v. 5, p. 1117 – 1133, 2013.

DORNFELD, C. B., Utilização de análises limnológicas, bioensaios de toxicidade e macroinvertebrados bentônicos para o diagnostic ambiental do reservatório de Salto Grande (Americana, SP). Dissertação de mestrado. Escola de Engenharia de São Carlos, Universidade de São Paulo, 2002.

DURBIN, J., KOOPMAN, S. J., Time series analysis by state space methods, Oxford University Press, 2012.

ECKERT, S., HÜSLER, F., LINIGER, F., HODEL, E., Trend analysis of MODIS NDVI time series for detecting land degradation and regeneration in Mongolia. Journal of Arid Environments, v. 113, p. 16 – 28, 2015. DOI: 10.1016/j.jaridenv.2014.09.001

EILANDER, D., ANNOR, F. O., IANNINI, L., GIESEN, N. V., Remotely sensed monitoring of small reservoir dynamics: a Bayesian approach. Remote Sensing, 6, 1191 – 1210, 2014. DOI: doi:1O.3390/rs6021191

FONSECA, M. F. Análise territorial do entorno do reservatório de Salto Grande (SP) com o uso de geoprocessamento: contribuições para usos múltiplos da água. Tese (Doutorado em Geografia) – Instituto de Geociências, Universidade Estadual de Campinas. Campinas, p. 266. 2013.

EASTMAN, J. R., FULK, M., Long sequence time series evaluation using standardized principal components. Photogrammetric Enginerring & Remote Sensing, v. 59, no. 8, p. 1307 – 1312, 1993.

FONSECA, M. F.; MATIAS, L. F. Análise do uso da terra e do componente clinográfico por meio de geoprocessamento: o entorno do Reservatório de Salto Grande – SP. Bol. Geogr., v. 32, n. 3, p. 48- 60, 2014. DOI: 10.4025/bolgeogr.v32i3.20888

FRANÇA, D. A.; FERREIRA, N. J.; FREITAS, S. R. Detecção de mudanças do uso e cobertura da terra em São José dos Campos – SP e vizinhanças e avaliação dos seus impactos no clima local. Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, INPE, p. 3905-3912, 2007.

FUSILLI, L., COLLINS, M. O., LANEVE G., PALOMBO, A., PIGNATTI, S., SANTINI, F., Assessment of the abnormal growth of floating macrophytes in Winam Gulf (Kenya) by using MODIS imagery time series. International Journal of Applied Earth Observation and Geoinformation, v. 20, p. 33 – 41, 2013. DOI: 10.1016/j.jag.2011.09.002

GALLACHER, K., MILLER, C., SCOTT, E. M., WILLOWS, R., POPE, L., DOUGLASS, J., Flow-directed PCA for monitoring networks. Environmetrics, v. 28, 2017. DOI: 10.1002/env.2434.

GIDEON, S., Estimating the dimension of a model. The Annals of Statistics, v. 6, p. 461 – 464, 1978.

GÓMEZ, C., WHITE, J. C., WULDER, M. A. Optical remotely sensed time series data for land cover classification: A review. ISPRS Journal of Photogrammetry and Remote Sensing, v. 116, p. 55 – 72, 2016. DOI: 10.1016/j.isprsjprs.2016.03.008

GRANGER, C. W. J., Investigating causal relations by econometric models and cross-spectral methods. Econometrica, v. 37, n. 3, 1969.

HANSEN, M. C., LOVELAND, T. R., A review of large area monitoring of land cover change using Landsat Data. Remote Sensing of Environment, 122, 66 – 74, 2012. DOI: https://doi.org/10.1016/j.rse.2011.08.024

HEROLD, M., ROBERTS, D. A., GARDNER, M. E., DENNISON, P. E., Spectrometry for urban area remote sensing – Developmente and analysis of a spectral library from 350 to 2400nm. Remote Sensing of Environment, v. 91, p. 304 – 319, 2004. DOI: 10.1016/j.rse.2004.02.013

HÖFLING, J. C., FERREIRA, L. I., RIBEIRO NETO, F. B., BRUNINI, A. P. C., Ecologia trófica do reservatório de Salto Grande, Americana, SP, Brasil. Bioikos, v. 14, p. 7 – 15, 2000.

JACQUIN, A., SHEEREN, D., LACOMBE, J. P.,Vegetation cover degradation assessment in Madagascar savanna based on trend analysis of MODIS NDVI time series. International Journal of Applied Earth Observation and Geoinformation, v. 12S, 2010. DOI: 10.1016/j.jag.2009.11.004

JAKUBAUSKAS, M. E.; PETERSON, D. L.; KASTENS, J. H.; LEGATES, D. R. Time Series Remote Sensing of Landscape-Vegetation Interactions in the Southern Great Plains. Photogrammetric Engineering & Remote Sensing**,** vol. 68, No. 10, p.1021-1030, 2002.

JAMALI, S., JÖNSSON, P., EKLUNDH, L., ARDÖ, J., SEAQUIST, J., Detecting changes in vegetation trends using time series segmentation. Remote Sensing of Environment, v. 156, p. 182 – 195, 2015. DOI: 10.1016/j.rse.2014.09.010

JENSEN, R., Remote Sensing of the Environment: An Earth Resource Perspective., Pearson, 2000.

JIA, K., LIANG, S., ZHANG, L., WEI, X., YAO, Y., XIE, X., Forest cover classification using Landsat ETM+ data and time series MODIS NDVI data. International Journal of Applied Eath Observation and Geoinformation, 33, 32 – 38, 2014. DOI: 10.1016/j.jag.2014.04.015

JOLLIFFE, I. T., CADIMA, J., Principal component analysis: a review and recent developments. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, v. 374, n. 2065, 2016. 10.1098/rsta.2015.0202

JOHNSON, R., WICHERN, D. W.,Applied multivariate statistical analysis. Sixth edition, Pearson Prentice Hall, Upper Saddle River, NJ, 2007.

JÖNSSON, P.; EKLUNDH, L. TIMESAT - a program for analyzing time-series of satellite sensor data. Computers & Geosciences, v. 30, n. 8, p. 833 – 845, 2004. DOI: 10.1016/j.cageo.2004.05.006

JONES, S. K., FREMIER, A. K., DeCLERCK, F. A., SMEDLEY, D., PIECK, A. O., & MULLIGAN, M. Big data and multiple methods for mapping small reservoirs: comparing accuracies for application in agricultural landscapes. Remote Sensing, v. 9, n. 1307, 2017. DOI: 10.3390/rs9121307.

KALNAY, E., KANAMITSU, M, KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G., WOOLLEN, J., ZHU, Y., LEETMAA, A., REYNOLDS, R., CHELLIAH, M., EBISUZAKI, W., HIGGINS, W., JANOWIAK, J., MO, K. C., ROPELEWSKI, C., WANG, J., JENNE, R. AND JOSEPH, D. (1996) The NCEP/NCAR 40-year reanalysis project, Bulletin of the American Meteorological Society v .77, p. 437-471

KIM, B.Y., LEE, L-T., JEE, J-B., Zo, I-S. Retrieval of outgoing longwave radiation at top-of-atmosphere using Himawari-8 AHI data. Remote Sensing of Environment, v. 204, p.498-508, 2018. DOI: 10.1016/j.rse.2017.10.006

KOTCHENOVA, S. Y., VERMOTE, E. F., MATARRESE, R., JR. KLEMM. F. J., Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. Applied Optics, v. 45, 2006. DOI: 10.1364/AO.45.006762

KRAEMER, G., CAMPS-VALLS, G., REICHSTEIN, M., MAHECHA, M. D., Summarizing the state of the terrestrial biosphere in few dimensions. Biogeosciences Discussions, in review. DOI: 10.5194/bg-2019-307

LACOUL, P., & FREEDMAN, B., Environmental influences on aquatic plants in freshwater ecosystems. Environmental Reviews, v. 14, p. 89 – 136, 2014 DOI: 10.1139/A06-001.

LACOUL, P., & FREEDMAN, B., Relationships between aquatic plants and environmental factors along a steep Himalayan altitudinal gradient, Aquatic Botany, v. 84, p. 3 – 16, 2006a. DOI: 10.1016/j.aquabot.2005.06.011.

LEBRETON, C., STELZER, K., BROCKMANN, C., BERTELS, L., PRINGLE, N., PAPERIN, M., DANNE, O., KNAEPS, E., RUDDICK, K., METH-103 – Cloud and cloud shadow masking of high and medium resolution optical sensors – na algorithm inter-comparison example for Landsat 8. Proceedings of ESA Living Planet Symposium, Prague, 2016.

LEE, E., HE, Y., ZHOU, M., LIANG, J., Potential feedback of recent vegetation changes on summer rainfall in the Sahel. Physical Geography, 2015. DOI: 10.1080/02723646.2015.1120139

LEITE, M. A., Variação espacial e temporal da taxa de sedimentação no reservatório de Salto Grande (Americana – SP) e sua influência sobre as características limnológicas do sistema. São Carlos (SP), Master Thesis, 1998.

LETTENMAIER, D. P., ALSDORF, D., DOZIER, J., HUFFMAN, G. J., PAN, M., WOOD, E. F., Inroads of remote sensing into hydrologic science during WRR era. Water Resour. Res, v. 51, 2015. DOI:10.1002/2015WR017616.

LI, J., YU, Q., TIAN, Y., BECKER, B. L., SIQUEIRA, P., TORBICK, N., Spatio-temporal variations of CDOM in shallow inland waters from a semi-analytical inversion of Landsat-8, 2018.

LIU, H., ZHENG, L., YIN, S., Multi-perspective analysis of vegetation cover changes and driving factors o long time series based on climate and terrain data in Hanjiang River Basin, China. Arabian Journal of Geosciences, v. 11. DOI:10.1007/s12517-018-3756-3.

LONDE, L. R. Comportamento espectral do fitoplâncton de um reservatório brasileiro eutrofizado – Ibitinga (SP), Tese de Doutorado: Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos – São Paulo, 2008.

LU, D. MAUSEL, P., BRONZÍDIO, E., MORAN, E. Change detection techniques. International Journal of Remote Sensing**,** v. 25, n. 12, p. 2365-2401, 2004. DOI: 10.1080/0143116031000139863

LUO, J. et al. Applying remote sensing techniques to monitoring seasonal and interannual changes of aquatic vegetation in taihu lake, China. Ecological Indicators, v. 60, p. 503–513, 2015.

MACEDO, C. R., Utilização de dados e técnicas de sensoriamento remoto no monitoramento da infestação por plantas aquáticas em reservatório de hidrelétrica. Faculdade de Ciências e Tecnologia Júlio de Mesquita Filho, UNESP. Presidente Prudente (SP). Graduation thesis, 2011. DOI: /10.17271/19800827722011118

MACHADO-MACHADO, E. A, NEETI, N., EASTMAN, J. R., CHEN, H., Implications of space-time orientation for principal component analysis of Earth observation image time series. Earth Sci Information, v. 4, p. 177 -124, 2011. DOI: 10.1007/s12145-011-0082-7.

MAHMOODZADEH, H. Digital change detection using remotely sensed data for monitoring green space destruction in Tabriz. Int. J. Environ. Res**.**, v. 1, p. 35-41, 2007.

MÄKELA, S., HUITU, E., & ARVOLA L., Spatial patterns in aquatic vegetation composition and environmental covariate along chains of lakes in the Kokemäenjoki watershed (S. Finland). Aquatic Botany, v. 80, p. 253 – 269. DOI: 10.1016/j.aquabot.2004.08.006, 2004.

MAKENZIE, F. T., Our changing planet: An introduction to Earth system science and global environmental change. 3rd edition, Prentice Hall, New Jersey, 2003.

MARTINS, D., COSTA, N. V., TERRA, M. A., MARCHI, S. R., Caracterização da comunidade de plantas aquáticas de dezoito reservatórios pertencentes a cinco bacias hidrográficas do estado de São Paulo, Planta Daninha, v. 26, p. 17 – 32., 2008. DOI: 10.1590/S0100-83582008000100003

MARTINS, D., MARCHI, S. R., COSTA, N. V., CARDOSO, L. A., RODRIGUES-COSTA, A. C., Levantamento de plantas aquáticas no reservatório de Salto Grande-SP, Brazil. Planta Daninha, v. 29, no. 1, 2011. DOI: 10.1590/S0100-83582011000100025

MARTINEZ, B., GILABERT, M. A., Vegetation dynamics from NDVI time series analysis using wavelet transform. Remote Sensing of Environment, v. 113, p. 1823 – 1842, 2009. DOI: 10.1016/j.rse.2009.04.016

MEANS, J. D., CAYAN, D. Precipitable water from GPS Zenith delays using North America regional reanalysis meteorology. American Meteorological Society, p. 485 – 495, 2013. DOI: 10.1175/JTECH-D-12-00064.1

MILTON, E.J. Principles of Field Spectroscopy. International Journal of Remote Sensing. 187, p. 1807-1827. 1987. DOI: 10.1080/01431168708954818

MORISSETTE, L., CHARTIER, S., The k-means clustering technique: general considerations and implementation in Mathematica. Tutorials in Quantitative Methods for Psychology, v. 9, p.15-24, 2013. DOI: 10.20982/tqmp.09.1.p015.

MYENI, R. B., HALL, F. G., SELLERS, P. J., MARSHAK, A. L. The interpretation of spectral vegetation indices. IEEE Transactions on Geoscience and Remote Sensing, v. 32, no. 2, 1995. DOI: 10.1109/36.377948

NEDERLANDS, K. (2000). Handbook for the meteorological Observation. Meteorologisch Instituut, KNMI. Available at: http://projects.knmi.nl/hawa/pdf/Handbook_H01_H06.pdf . Last visited in February, 2020.

NEETI, N., EASTMAN, J. R., Novel approaches in extended principal components analysis to compare spatio-temporal patterns among multiple image time series. Remote Sensing of Environment v. 148, p. 84 – 96, 2014. DOI: 10.1016/j.rse.2014.03.015

NORRULASHIKIN, S. M., YUSOF, F., KANE, I. L., Instantaneous causality approach to meteorological variables bond. Advances in Industrial and Applied Mathematics, v. 1750, AIP Conf. Proc. 1750, 060010-1–060010-8. 2016. DOI: 10.1063/1.4954615

ONG´ANG´A, O., MUNYIRWA, K., Ecological threat to aquatic and human life in and around Lake Victoria. Seminar for representatives of Ministry of health and public health management training institutions, Entebbe Uganda, 1997.

PHAM, D. T., DIMOV, S., NGUEN, C. D., Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science, v. 219, p. 103-119, 2005. DOI: 10.1243/095440605X8298.

PERMATASARI, P. A., FATIKHUNNADA, A., LIYANTONO, SETIAWAN, Y., SYARTINILIA, NURDIANA, A., Analysis of agricultural land use changes in Jombang Regency, East Jav, Indosenia using BFAST method. Procedia Environmental Sicence, v. 33, p. 27 – 35, 2016 DOI: 10.1016/j.proenv.2016.03.053

QIU, S., HE, B., ZHU, Z., LIAO, Z., QUAN, X., Improving Fmask cloud and shadow detection in mountainous area for Landsat 4-8 images. Remote Sensing of Environment, v. 199, p. 107-119, 2017. DOI : 10.1016/j.rse.2017.07.002

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria, 2016.

REN, Z., QU, X., ZHANG, M., YU, Y., PENG, W., Distinct bacterial communities in wet and dry seasons during a seasonal water level fluctuation in the largest freshwater lake (Poyang Lake in China). Frontier in Microbiology, 2019. DOI: 10.3389/fmicb.2019.01167/full

RICHARDS, J., JIA, X., Remote sensing digital image analysis. New York: Springer, ed. 5, 2013.

RIESER, D., KUHN, M., PAIL, R., ANJASMARA, I. M., AWANGE, J., Relation between GRACE-derived surface mass variations and precipitation over Australia. Australian Journal of Earth Sciences, v. 7, p. 887 – 900, 2010. DOI: 10.1080/08120099.2010.512645

RODRIGUES, L. N., SANO, E. E., STEENHUIS, T. S., PASSO, D. (2012). Estimation of small reservoir storage capacities with remote sensing in the Brazilian savannah region. Water Resources Management, v. 26, p. 873-882, 2012.

ROSSUM, G. V. Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, 1995.

SALLES, M. A., CANZIANI, P. O., AND COMPAGNUCCI, R. H., The spatial and temporal behavior of the lower stratospheric temperature over the southern hemisphere: the MSU view. PART II: spatial behaviour. International Journal of Climatology, v. 21, p. 439 – 454, 2011. DOI: 10.1002/joc.607.

SÁ, E. A. S., MOURA, C. N., PADILHA, V. L., CAMPOS, C. G. C., Trends in daily precipitation in highlands region of Santa Catarina, southern Brazil. Ambiente & Água – na interdisciplinary Journal of Applied Science, v. 13, 2018. DOI: 10.4136/ambi-agua.2149.

SANTOS, A. G., Variação da pressão em Tahiti e sua relação com a precipitação no Brasil com ênfase o leste do nordeste. Universidade Federal de Alagoas (UFAL). Maceió. Dissertação, 2011.

SANTOS JUNIOR, A., COSTACURTA, M. B., Dinâmica da composição e cobertura de espécies de macrófitas aquáticas e a escolha de indicadores de impacto ambiental em um rio com ecoturismo. Ambiência, v. 7, p. 535 – 550, 2001.

SCHOWENGERDT, R. A. Remote Sensing: models and methods for image processing. v. 3. London: Academic Press, 2006

SHEKEDE, M. D., KUSANGAYA, S., SCHMIDT, K., Spatio-temporal variations of aquatic weeds abundance and coverage in Lake Chivero, Zimbabwe. Physics and Chemistry of the Earth, Parts A/B/C, v. 33, p. 714 – 721, 2008. DOI: /10.1016/j.pce.2008.06.052

SILVA, T. S. F., COSTA, M. P. F., MELACK, J. M., Spatial and temporal variability of macrophyte cover and productivity in the eastern Amazon floodplain: a remote sensing approach. Remote Sensing of Environment v. 114, p. 1998 – 2010, 2010. DOI: 10.1016/j.rse.2010.04.007

TAN, P-N., STEINBACH, M., KARPATNE, A., KUMAR, V., Introduction to Data Mining. 2nd Edition, Pearson, 2019.

TENA, A., VERICAT, D., GONZALO, L. E., BATALLA, R. J., Spatial and temporal dynamics of macrophyte cover in a large regulated river. Journal of Environmental Management, v. 202, p. 379 – 391, 2017. DOI: /10.1016/j.jenvman.2016.11.034

TIAN, F., FENSHOLT, R., VERBESSELT, J., GROGAM, K., HORION, S., WANG, Y., Evaluation temporal consistency of long-term global NDVI datasets for trend analysis. Remote Sensing of Environment, v. 163, p. 326 – 340, 2015. DOI: /10.1016/j.rse.2015.03.031

TOLPEKIN, V. A., STEIN, A. The core of GISsciente: a system-based approach, Enschede: University of Twente Faculty of Geo-Information and Earth Observation (ITC), 2013.

TUBELIS, A., NASCIMENTO, F. J. L., Meteorologia descritiva; fundamentos e aplicações brasileiras. São Paulo: Nobel, 1980.

USGS, U. S. Geological Survey, Department of the Interior. Landsat 7 (L7) Data Users Handbook, Version 1.0, 2018. Available at: https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1927_L7_Data_Users_Handbook-v1.pdf. Last visited on September, 2019.

USGS, U. S. Geological Survey, Department of the Interior. Landsat 8 (L8) Data Users Handbook, Version 4.0, 2019. Available at: https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1574_L8_Data_Users_Handbook_v4.pdf. Last visited on September, 2019.

VERBESSELT, J. HYNDMAN, R., ZEILEIS, A., CULVENOR, D., Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. Remote Sensing of Environment, 2010. DOI: 10.1016/j.rse.2010.08.003

VERBESSELT, J., HYNDMAN, R., NEWNHAM, G., CULVENOR, D., Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment, v. 114, no. 1, 2010b. DOI: 10.1016/j.rse.2009.08.014

VERMOTE, E., JUSTICE, C., CLAVERIE, M., & FRANCH, B., Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment, v. 185, p. 46-56, 2016.

VERSTRAETE, M. M., PINTY, B., Designing optimal spectral indices for remote sensing applications. IEEE Transactions on Geosciente and Remote Sensing, v. 34, no. 5, 1996. DOI: 10.1109/36.536541

VILLA, P., PINARDIA, M., BOLPAGNIB, R., GILLIERC, J. M., ZINKED, F. N., BRESCIANI, M. Assessing macrophyte seasonal dynamics using dense time series of medium
resolution satellite data. *Remote Sensing of Environment*, 216 (2018) 230–244. https://doi.org/10.1016/j.rse.2018.06.048, 2018.

WATTS, L., LAFFAN, S. W., Sensitivity of the BFAST algorithm to MODIS satellite and vegetation index. 20th International Congress on Modelling and Simulation, Adelaide, Australia, 2013.

WATTS, L., LAFFAN, S. W., Effectiveness of BFAST algorithm for detecting vegetation response patterns in a semi-arid region. Remote Sensing of Environment v. 154, p. 234 – 245, 2014. DOI: 10.1016/j.rse.2014.08.023

WESTLAKE, D. F., Macrophytes, In: River Ecology, Studies in Ecology, v. 2, Blackwell Scientific Publications, Oxford, UK, 1975.

WIEMANN, S., BERNARD, L., Spatial data fusion in spatial data infrastructures using linked data. International Journal of Geoinformation Science, p. 613-636, 2015. DOI: 10.1080/13658816.2015.1084420

WILBERFORCE, N., Climate change detection in Central Africa. Faculty of Environmental Science and Process Engineering, Brandenburg University of Technology Cottbus, Germany. Doctoral Thesis, 2007.

WILLIAMS, D. M., Time series analysis of vegetation dynamics and burn scar mapping at smoky hill air national guard range,, Kansas using moderate resolution satellite imagery.

Department of Geography, College of Arts and Sciences, Kansas State University. Master Thesis, 2013.

ZANATA, L. H., ESPÍNDOLA., E. L. G., Longitudinal processes in Salto Grande reservoir (Americana, SP, Brazil) and its influence in the formation of compartment system. Braz. J. Biol, v. 62, p. 347 – 361, 2002. DOI: 10.1590/S1519-69842002000200019

ZEILEIS, A., KLEIBER, C., KRÄMER, W., HORNIK, K., Testing and dating of structural changes in practice. Computacional Statistics and Data Analysis, v. 49, p. 109 – 123, 2003. DOI: 10.1016/S0167-9473(03)00030-6

ZEILEIS, A., KLEIBER, C., Validating multiple structural changes models – a case study. Journal of Applied Econometrics, p. 685 – 690, 2005. DOI: 10.1016/S0167-9473(03)00030-6

ZHA, Y., GAO, J., NI, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. International Journao of Remote Sensing, v. 24, p. 583 – 594, 2003. DOI: 10.1080/01431160304987

ZHANG, S., FOERSTER, S., MEDEIROS, P., ARAÚJO, J. C., MOTAGH, M., WASKE, B., Bathymetric survey of water reservoirs in north-eastern Brazil based on TanDEM-X satellite data. Science of the Total Environment, v. 571, p. 575 – 593, 2016. DOI: 10.1016/j.scitotenv.2016.07.024

ZHANG, S., FOERSTER, S., MEDEIROS, P., ARAÚJO, J. C., WASKE, B., Effective water surface mapping in macrophyte-covered reservoirs in NE Brazil based on TerraSAR-X time series. Int J. Appl. Earth Obs Geoinformation, v.. 69, p. 41 – 55, 2018. DOI: /10.1016/j.jag.2018.02.014

ZHANG L., YI J., Management methods of spatial data based on PostGIS. Second Pacific-Asia Conference on Circuits, Communications and Systems (PACCS), 2010. DOI: 10.1109/PACCS.2010.5626962

ZHANG, K., FU, R. Relationships between outgoing longwave radiation and diabatic heating in reanalyses. Climate Dynamics, 2016. DOI: 10.1007/s00382-016-3501-0

ZHU, Z., WOODCOCK, C. E., Automate cloud, cloud shadow, and snow detection in multiemporel Landsat data: an algorithm designed specifically for monitoring land cover change. Remote Sense of Environment, v. 152, p. 217 – 234, 2014. DOI: 10.1016/j.rse.2014.06.012

ZHU, Z., WANG, S., WOODCOCK, C. E. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. Remote Sense of Environment, v. 159, p. 269-277, 2015. DOI: 10.1016/j.rse.2014.12.014